DOCUMENT RESUME

ED 361 465 UD 029 499

TITLE	Reinforcing the Promise, Reforming the Paradigm. Report of the Advisory Committee on Testing in Chapter 1.
INSTITUTION	Office of Elementary and Secondary Education (ED), Washington, DC. Compensatory Education Programs.
REPORT NO	ED/OPP-93-22
PUB DATE	May 93
NOTE	74p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Achievement Tests; Advisory Committees; *Compensatory Education; Disadvantaged Youth; Educational Assessment; Educationally Disadvantaged; *Educational Policy; Elementary Secondary Education; *Federal Programs; Guidelines; *Program Evaluation; Remedial Instruction; Standardized Tests; Student Evaluation; *Test Use
IDENTIFIERS	*Education Consolidation Improvement Act Chapter 1

ABSTRACT

The Advisory Committee on Testing in Chapter 1 evaluated current testing policies in the Chapter 1 program and offered the U.S. Department of Education advice on improvements and alternatives to the current system. After analyzing existing testing procedures, the committee concluded that Chapter 1's overreliance on a single testing method, aggregated gain scores on standardized norm-referenced tests, does not provide adequate information by which to judge the progress of students, the quality of the school-level program, or the effectiveness of the national program. The Committee recommended that: (1) the Federal Government periodically evaluate the national effectiveness of Chapter 1 by evaluating the achievement of a representative sample of eligible student beginning in grade 3; (2) states should develop and implement several types of standards; (3) local education agencies should assess the progress of participant children using state-developed assessment methods; (4) teachers should assess students using a variety of methods; (5) school districts should use multiple indicators; (6) assessment should be compatible with schoolwide service delivery; (7) a transition period of 5 years should be allowed; and (8) funds should be set aside for staff development. A copy of the Committee charter is provided. (Contains 6 references.) (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 361 465

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: UD
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. TM

Reinforcing the Promise, Reforming the Paradigm

Report of the Advisory Committee on Testing in Chapter 1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. Department of Education

BEST COPY AVAILABLE

THE ADVISORY COMMITTEE ON TESTING IN CHAPTER 1

Eva L. Baker
Center for the Study of Evaluation
University of California, Los Angeles

Edward A. De Avila
Linguametrics Group
Oakland, California

Robert B. Frary
Office of Measurement and Research Services
Virginia Polytechnic Institute and State University

James W. Guthrie
School of Education
University of California, Berkeley

Althea Hall
Jersey City State College
State University of New Jersey

Sharon Lynn Kagan
Bush Center in Child Development and Social Policy
Yale University

Joy G. Korologos
Vice Chair, Management Society
Brigham Young University

Brenda Joyce Lee
Principal, Glenallan Elementary School
Montgomery County (Maryland) Public Schools

Milton D. Matthews
Director, Division of Compensatory Education
Mississippi Department of Education

Rae Ellen McKee
Chapter 1 Teacher
Hampshire County (West Virginia) Schools

Vincent E. Reed
Vice President
The Washington Post

Thomas A. Romberg
Director, National Center for Research in Mathematical Sciences Education
University of Wisconsin--Madison

Diane J. Sawyer
Office of Dyslexic Studies
Middle Tennessee State University

Ramsay W. Selden
State Education Assessment Center
Council of Chief State School Officers

Robert S. Stephenson
Office of Educational Accountability
Dade County (Florida) Public Schools
(retired)

Karen Kring Wixson
School of Education
University of Michigan

REINFORCING THE PROMISE, REFORMING THE PARADIGM

Report of the Advisory Committee on Testing in Chapter 1

May 1993

CONTENTS

PREFACE i

EXECUTIVE SUMMARY vi

REINFORCING THE PROMISE: THE CASE FOR REFORM OF CHAPTER 1
Testing. 1
Introduction 1
New Challenges for Chapter 1 2
The Role of Testing in Chapter 1 4
The Consequences of Chapter 1 Testing Procedures 9
Failure to Employ Multiple Measures 10
Inappropriate Uses 11
Impediments to Appropriate Curriculum and Instruction 12
Failure to Assess Special Populations Appropriately 13
Questionable Validity and Utility of Results 14

REFORMING THE PARADIGM: RECOMMENDATIONS FOR CHAPTER 1 TESTING 16
The Purpose of the Advisory Committee 16
Guiding Principles 16
Recommendations 20
National Accountability 20
State Accountability 23
Local Accountability 30
Instructional Feedback and Diagnosis 31
Student Eligibility and Identification 32
Schoolwide Projects 34
Transition Period 35
Federal Government Development Assistance 38

CONCLUSION 39

VIEWS FROM INDIVIDUAL COMMITTEE MEMBERS 40

APPENDICES 45
APPENDIX A A-1
APPENDIX B B-1

ACKNOWLEDGMENTS

PREFACE

This is a report about testing in Chapter 1. This is the popular label for the federally funded program to assist educationally disadvantaged students. The program was initially authorized in 1965 with enactment of the Elementary and Secondary Education Act. It has subsequently been reauthorized or significantly amended by Congress and the President on eight instances and currently is in its 27th operational year in America's schools.

We wish to reinforce strongly the importance of Chapter 1 for America's schools and for our entire society. Every fact, every conclusion, every recommendation in this report is aimed squarely at strengthening Chapter 1. The need for this federal financial assistance was great at the time of the program's initial enactment more than a quarter-century ago. In the intervening period the need has only become greater.

We also seek to stimulate change, to provoke a careful and practical reconsideration of the premises currently undergirding Chapter 1 testing and measurement. In our judgment, *now is the time to recalibrate the balance in Chapter 1 between the instructional needs of students and teachers and the program evaluation and accountability needs of the policy system.*

Chapter 1 and American education have evolved substantially since the program's 1965 inception. A quarter-century ago, Chapter 1's purposes were less well understood, our understanding of the manner in which organizations respond to regulations and financial inducements was less sure, and our knowledge of the fundamental components of instruction was less developed. In the face of such daunting uncertainty, an intense regulatory and financial accountability overlay seemed necessary. Ensuring that state and local governments complied with and spent their Chapter 1 funds for purposes defined by the federal mandate seemed appropriate.

Today it is possible to draw upon more than 25 years of Chapter 1 administrative experience, new understanding of organizational dynamics, new knowledge regarding the nature of human learning and growth, and far more sophisticated efforts in testing and measurement. Moreover, Chapter 1 has itself evolved into a principal component of American education. Its purposes and procedures are now widely accepted and well understood throughout our nation's states and school systems.

While the need for Chapter 1 has grown, the educational and organizational context in which it operates has changed. In order to extend its future effectiveness, Chapter 1--particularly those components connected with testing and measurement--must change too. The time has now arrived to define a new policy balance. Chapter 1 testing and evaluation regulations should become *less* concerned with large scale evaluation and ensuring state and school procedural adherence and *more* oriented toward enhancing effective classroom instruction and elevating student achievement.

Put even more simply, the time is now appropriate for Chapter 1 testing to concentrate more on promoting student learning and less on measuring regulatory compliance.

We are eager that our message not be misunderstood. The new balance we seek between, on the one hand, federal government regulatory compliance and program accountability and, on the other hand, instructional effectiveness, local school creativity, and professional educator responsibility is not the product of an historic vacuum. We have learned well from the lessons of past practice. We are mindful of prior abuses wherein Chapter 1 resources were used for purposes other than those for which they were principally intended. We can imagine circumstances in which this unfortunate history could repeat itself.

Thus, we are *not* advocating a "hands off," "trust me totally," policy environment in which state and local educators are free to undertake any activity of their choosing in the name of Chapter 1. We believe that there is now and will continue to be a practical need for ensuring that Chapter 1 students are well served and the public's interests fulfilled. Thus, our recommendations, while suggesting a new operational paradigm, nevertheless continue to provide for program evaluation and operational accountability.

The principal change we advocate is a shift away from procedural compliance and toward a concentration on instruction and student learning outcomes. Practically this translates to the use of tests which are operationally linked to instructional objectives for students. Chapter 1 tests should be tightly tied to what students are expected to know and be able to do. In effect, tests should be so fundamentally integrated into regular instructional activities that students would frequently not be able to distinguish assessment from the regular flow of teaching in a classroom or school. Also, tests should be designed with careful consideration of their appropriateness for the age, grade level, and developmental stages of the students for whom they were intended. Finally, assessment should be sufficiently linked to instructional purposes that a school's professional staff could rely on test results to inform them of the degree to which their instructional strategies were succeeding, both with individual students and with groups of students.

The research understandings, organizational agreements, and technological developments necessary to attain the above-described ends are substantial. Consequently, one cannot expect an immediate transition toward an outcome orientation for all of Chapter 1 and its state and local components. As wise and well-intentioned as executive and legislative branch officials may be, it will almost assuredly take a transition period--perhaps as long as five years--to strike a creative balance between the present financial and procedural regulatory format and a new, badly needed, student achievement orientation.

During this transition, proponents of change will no doubt at times become frustrated, and advocates of the status quo will no doubt feel vindicated. Nevertheless, even if time-consuming, we believe that the transition to a new paradigm must eventually occur, lest

Chapter 1 lose its present effectiveness and fail to meet future challenges.

This report has four sections. The first explains that the nation's need for Chapter 1 is, if anything, greater today than ever before. Here we also stress that while our committee charge is centered on testing, this activity is itself so central to instruction and schooling that of necessity our report touches on additional operational features of schooling.

Second, we explain the testing and evaluation dynamics that presently characterize Chapter 1 operations and, in the process, describe several of the dysfunctional features that have evolved over time from the current regulatory outlook and over-emphasis upon lower order school skills. A third section describes the student performance outcome orientation, which we believe should characterize Chapter 1 in the future. A final section presents the recommendations on testing that we believe are at the heart of the shift we seek.

Throughout the report, we periodically call the readers' attention to important related matters, such as the application of the new Chapter 1 assessment strategy to language minority and handicapped students, which we believe are deserving of particular attention.

Implementation of the new paradigm will require a sustained high level of leadership and cooperation from the United States Department of Education, state education departments, local school districts and schools, education professionals and professional associations, researchers, and development specialists as well as a long list of related agents and agencies. In addition, funding, either new money or redirected funds, will have to be made available from federal sources for crucial activities such as the development of curriculum frameworks and student performance expectations, technical assistance to states and local schools districts, and the inservice preparation of teachers.

Our advisory committee is composed of concerned members of the general public, professional educators, private sector executives, professors, and policymakers. Our discussions have drawn on the expert knowledge of a wide range of school practitioners, academic experts, test manufacturers, and state and federal education officials. Deliberations have been open, candid, complicated, and intense. Regardless of individual members' points of view, however, discussions and decisions have uniformly been motivated by a desire to assist students most in need, to enhance the effectiveness of America's schools, and to act in the public's best long term interests.

The complexity of the testing topic and intensity of the points of view have led, inevitably perhaps, to differences in opinion among committee members about recommendations. These differences cluster around three dimensions: (a) the extent to which testing for accountability purposes should extend to Chapter 1 students below grade 3; (b) the relative utility of norm-referenced tests; and (c) the extent to which various Chapter 1 testing procedures sufficiently take into account the differences between language minority and disabled students and all others. The views of a few committee members are appended so that readers can more fully grasp the complexity and subtlety of these issues.

However, individual and dissenting views should not detract from the substantial agreement that the majority of the committee reached about fundamental items. Specifically, the committee was in agreement that--

Chapter 1 testing should become aligned with learning goals and expected student outcomes.

More than one mode of testing will be necessary to serve the diverse purposes connected with instruction and accountability.

The United States Department of Education should cooperate with states in developing the new procedures and tests that will be necessary to achieve the recommendations contained in this report.

Federal resources should be allocated to assist in the development and implementation efforts that this report's recommendations entail.

This report is intended for a broad spectrum of potential consumers. Paradoxically, its principal audience is composed of political officials, the Secretary of Education, other officers of the administration, and members of Congress. These are the actors who initially are positioned to receive and respond to our recommendations.

The paradox resides in the fact that this principal audience is understandably political while the committee is itself quite apolitical. Deliberations and decisions among committee members have been undertaken in a context totally devoid of partisan considerations.

This report is also intended for America's professional educators, teachers, administrators, and hosts of others who regularly strive to enhance the learning and fulfillment of individual students, their parents, and the larger society.

Another audience for this report consists of the thousands of parents, private citizens, professionals, and private sector business officials who routinely care about and cater to the needs of students and schools.

Finally, we as a committee are aware of the intense national attention on the 1993 Chapter 1 reauthorization activities. We are also aware that this report is only one of many that will occupy the attention of policymakers and other public officials. We appreciate the complexity and competing views that reauthorization will evoke. We admire the thoughtfulness of those who have preceded us in the historic efforts to forge Chapter 1 into an ever more effective instrument for the education of students, and we are pleased to have been asked to contribute to the current effort to improve it.

James W. Guthrie, Chair

EXECUTIVE SUMMARY

As the largest federal school aid program, the Chapter 1 program for disadvantaged children is an influential force in American education. Testing is one particularly strong area of influence. Millions of school children take standardized tests every year because of Chapter 1 testing requirements. Standardized tests, primarily the norm-referenced kind, are used to help determine which children should be served, to assess how much program participants are learning, and to evaluate whether the program is effective in individual schools and for the nation as a whole. Most of these functions relate in some way to the goal of "accountability"--ensuring that Chapter 1 funds are used well, to help improve the achievement of disadvantaged children.

Few would disagree that there should be strong accountability for Chapter 1, and that this accountability should include an appraisal of student progress. But the world has changed considerably since the current Chapter 1 testing system was put in place. Knowledge about teaching and learning has expanded. New approaches to testing have been piloted or implemented. Demands for higher educational standards for all students have emerged. Consequently, new questions have arisen about whether the current Chapter 1 testing requirements are keeping pace.

The Advisory Committee on Testing in Chapter 1 was established to advise the U.S. Department of Education on improvements or alternatives to the current testing system. After analyzing existing testing procedures, the committee has concluded that Chapter 1's overreliance on a single testing method--aggregated gain scores on standardized, norm-referenced tests--does not provide adequate information by which to judge the progress of students, the quality of the school-level program, or the effectiveness of the national program. Rather, the committee has concluded, the current testing requirements tend to reinforce some of the more ineffective or outmoded approaches to teaching disadvantaged children, such as drilling students on low-level basic skills or giving them less challenging subject matter than their peers receive. The weaknesses of the current system have become more pronounced since enactment of the 1988 Amendments to Chapter 1, which raised the stakes attached to Chapter 1 testing by requiring schools that showed insufficient test score gains to engage in a program improvement process.

The committee therefore recommends a new approach to Chapter 1 assessment and accountability, based on five important principles.

> *Principle #1: Chapter 1 should continue to have strong accountability, but the balance should shift to emphasize how well students are learning and how effectively they are being taught.*

The current emphasis in Chapter 1 testing is on compliance with evaluation procedures and mandates. After 27 years of experience with Chapter 1, states and local districts understand and respect its basic goals and are ready for a new degree of flexibility

and creativity in assessment. In exchange, however, they should be able to demonstrate that Chapter 1 children are progressing toward ambitious expectations for learning, and that schools are providing high-quality instruction. To make these determinations, states and districts will need to use multiple measures aligned more closely with the types of student learning outcomes being sought.

> *Principle #2: Chapter 1 testing should no longer be an independent system but should be linked with the education reforms that states and school districts are undertaking for* all *children.*

Right now several professional associations and study groups, including the panels following progress toward the National Education Goals, are developing high, voluntary national standards for what American students should know and be able to do in key subjects. Chapter 1 students should be prepared to reach those standards, or whatever high expectations states set for all children. Toward this end, the committee proposes that Chapter 1 accountability be based on assessments that are aligned with high standards for the content all children should learn and the performance all children should attain in reading, writing, oral language, mathematics, and to the extent possible the other subjects in the National Education Goals.

> *Principle #3: National Chapter 1 evaluation should be decoupled from state, local, and classroom assessment functions.*

It is the need for aggregated national data that has driven much of the reliance on a single form of testing. Meeting national evaluation needs with special assessment would give states, districts, schools, and teachers greater flexibility to design Chapter 1 accountability approaches that are better aligned with education and assessment reforms for all children.

> *Principle #4: The diverse purposes of assessment in Chapter 1 should be met with multiple methods of assessment.*

Chapter 1 should require multiple assessment methods at all levels of government, including performance assessments that require students to undertake an action or create a product demonstrating their knowledge or skills.

> *Principle #5: Chapter 1 assessment should recognize the different developmental stages of children.*

The committee supports the concept of early intervention but recognizes that care must be taken in assessing young children, defined in this report as children below grade 3. Therefore, the committee recommends different assessment strategies for children of different ages and grade levels.

How can these principles be implemented? The committee offers eight specific

recommendations, five that pertain to the major functions of Chapter 1 testing at the national, state, local, classroom, and student levels, followed by three more that cut across all levels and functions.

> *Recommendation #1: The federal government should periodically evaluate the national effectiveness of Chapter 1 using a NAEP-like assessment that evaluates the achievement of a representative sample of Chapter 1-eligible students beginning in grade 3; collects background information on Chapter 1 students and programs; and analyzes the long-term effects of Chapter 1 participation. Special procedures should be used to ensure national accountability below grade 3.*

The federal government should design and implement a national assessment to meet national accountability needs, based on the sampling, quality control, and other technical procedures used by the National Assessment for Educational Progress (NAEP). It is not necessary to test every Chapter 1 child every year to obtain a reliable national picture of Chapter 1's effectiveness. In fact, the current system of aggregating millions of test scores upward through the district, state, and federal levels is an inefficient and sometimes imprecise way of doing a national evaluation. Through a NAEP-like sampling approach, a national assessment should evaluate the achievement of a representative sample of Chapter 1-eligible children in reading, writing, oral language, mathematics, science, history and geography. The assessment could be conducted on a multi-year cycle, rather than annually, and could be implemented in selected grades, beginning with grade 3. The assessment should also collect background information about Chapter 1 students and programs and analyze the long-term effects of Chapter 1 participation. A well-designed assessment of this nature could provide Congress with better information than it receives now.

Children in prekindergarten through grade 2 should not be included in the national assessment. There should, however, be special national studies at grade 2, using performance-based assessments that meet other strict criteria to ensure appropriate, sensitive assessment of young children. In addition, the Secretary should review data on program delivery for prekindergarten through grade 1.

> *Recommendation #2: States should develop and implement: (a) content and performance standards that will address what Chapter 1 students should know and be able to do, which should be the same as standards for all students; (b) Chapter 1 assessment methods aligned with these content and performance standards; (c) delivery standards for effective Chapter 1 programs; and (d) procedures for local reporting of student assessment results and state monitoring of Chapter 1 program delivery and effectiveness.*

States should assume a stronger leadership role in Chapter 1 assessment and accountability. Therefore, states would be the linchpin of the new paradigm for accountability. As a first phase, states should develop and implement high standards for Chapter 1 content and student performance that are the same as state standards for all

children. As part of this process, states should consider whatever voluntary national standards exist for key subjects, as they become available. In the next phase, states should design and implement a system of multiple assessments for Chapter 1, including alternatives to conventional standardized tests, that are aligned with the state content and performance standards. The standards and assessments resulting from this process would be submitted to the U.S. Secretary of Education for approval and would guide Chapter 1 assessment and accountability at the state and local levels.

Because it is not fair to expect students to perform at a certain level without also ensuring that they receive meaningful opportunities to learn, the committee also recommends that states develop delivery standards addressing the elements, practices, and inputs that contribute to a high-quality Chapter 1 program. States and local districts would use these delivery standards as a basis for evaluating program quality at the school and classroom levels. As a final component of the state role, the committee recommends that states develop procedures for local reporting of Chapter 1 assessment results and for state monitoring of program effectiveness, which should include classroom observations of program delivery.

For state and local accountability purposes, programs at the prekindergarten and kindergarten levels would be assessed on the basis of delivery standards only. At grades 1 and/or 2, there would be an assessment using both delivery standards and some student content and performance standards, provided that assessments were performance based and developmentally appropriate.

> *Recommendation #3: To ensure accountability for federal funds, local education agencies should assess the progress of Chapter 1 children in grades 3 through 12 using state-developed assessment methods (or locally adapted ones where states permit) tied to content and performance standards for all children. Different procedures should be used to assess programs for prekindergarten and kindergarten and for grades 1 and 2.*

Local accountability should be closely linked to state accountability. At the option of the state, local school districts could be allowed to modify state standards and assessments or to develop their own standards and assessments that met similar criteria. The accountability system of content and performance standards, delivery standards, assessments, and monitoring procedures could form a basis for determining the effectiveness of programs at the school level, as well as the progress of individual students. The committee stresses, however, that these determinations should be based on multiple measures.

> *Recommendation #4: Teachers should assess Chapter 1 student learning and diagnose student needs using continuous, intensive, and varied methods and should use the feedback from these assessments to plan and adjust instruction.*

Teachers need a range of information to monitor student learning, diagnose student needs, and inform their own teaching. This instructional feedback function of Chapter 1 testing is one of the most important functions, but is among the most neglected by the current testing requirements. The aim of this recommendation is to give teachers the encouragement and the tools they need to incorporate good assessment practices into their everyday classroom operations. This function of assessment should be controlled by teachers. The results of state and local accountability assessments would be just one source of feedback; teachers would determine what others were needed, which could include informed teacher judgment, classroom observations, performance assessment, and other assessments.

> *Recommendation #5: School districts should use multiple indicators to identify students with the greatest needs for Chapter 1 services, including special procedures for young children. Appropriate methods should be used to identify and select limited-English-proficient students and special education students who are educationally disadvantaged for inclusion in Chapter 1 programs.*

Children at the prekindergarten through grade 2 level should be selected for Chapter 1 primarily on the basis on poverty, plus consideration of other factors that may place children at educational risk and informed teacher judgment. Special care should also be taken to include limited-English-proficient (LEP) children and special education children in Chapter 1 programs and to assess them appropriately. For LEP children, assessments for both accountability and eligibility purposes should include an assessment of oral language.

> *Recommendation #6: Chapter 1 assessment should be compatible with schoolwide approaches to service delivery, as well as other effective models for Chapter 1 instruction.*

The committee recommends the schoolwide project approach, in which Chapter 1 funds are used to upgrade instruction for all children attending the highest-poverty schools, as a highly desirable option for Chapter 1 services. When well implemented, the schoolwide project approach helps link Chapter 1 assessment with educational reforms for all children.

> *Recommendation #7. The five years following the forthcoming reauthorization of Chapter 1 should serve as a transition period, with new elements phased in as they become ready. By the end of the five years, state standards, assessments, and procedures for Chapter 1 accountability should be in place, and a national assessment for Chapter 1 should be operational. During this transition, Chapter 1 assessment should operate according to state transition plans, approved by the Secretary, that include multiple assessment methods.*

A great deal of research, development, training, consensus building, and other work will need to be done to bring about a shift of the magnitude proposed in this report. A five-year transition period would give states time to develop and put in place standards, assessments, and procedures. The federal government would develop and begin to

implement the national accountability assessment, and would provide staff development, research, and technical assistance to state and local agencies. By the end of five years, all elements of the new accountability system should be in place.

Upon enactment of new amendments to Chapter 1, the committee recommends that the current system of nationally aggregated norm-referenced test data be discontinued, and that states immediately develop transition plans for ensuring Chapter 1 accountability during the interim five-year period, until the new system is ready.

> *Recommendation #8: To help teachers, state and local administrators, and other key professionals implement these recommendations, the Chapter 1 legislation should include a funding set-aside for staff development related to assessment and performance standards. The federal government should also support a national effort to expand and refine the knowledge base about assessment and standards.*

Staff development is critical to the success of all the committee's recommendations. Also vital is additional research and development in such areas as alternative assessment, new assessments linked to standards, assessment of young children, and the role of technology in assessment.

GLOSSARY OF CHAPTER 1 TESTING TERMS

Standardized tests are administered and scored under conditions uniform to all students. Standardization is a generic concept that can apply to any testing format--from multiple choice to written essays to performance estimates. Standardization makes scores comparable and assures, to the extent possible, that test takers have equal chances to demonstrate what they know.

Norm-referenced tests are standardized tests that compare a student's performance against that of other test-takers. Norms are obtained by administering the test to a given population (the norm group) and then calculating means, standard deviations, standardized scores, and percentiles.

Testing procedures, as used in this report, refer to the totality of requirements in Chapter 1 law, regulations, and policy guidelines that affect the type, frequency, use, and reporting of tests and test results.

Assessment, in an educational context, is a generic term that covers the range of processes used to determine or estimate what students know and can do and how much they have learned. Assessment can include tests, student learning demonstrations, teacher observations, professional judgments, and other indicators such as attendance, graduation rates, and surveys.

Evaluation, in the context of Chapter 1, refers to the processes used to determine how much academic progress students have made and, on a broader level, how well Chapter 1 programs are operating. Chapter 1 law and regulations prescribe procedures for local and state evaluations. Consequently, evaluation generally refers to a broader, more formal process than assessment or testing.

A *test score* is an estimate of what the test-taker knows or can do based on a sample of items on a test. Because they are based on samples of behavior, test scores must be interpreted carefully.

Nationally normed tests use a norming group that is a representative sample of a component of the national population (e.g., third-graders).

A *normal curve equivalent* (NCE) is a statistic that is based on a scale similar to a percentile rank but that permits aggregation of results. In technical terms, an NCE is a standard score resulting from the division of the normal curve into 99 equal units; the 50th percentile (an NCE of 50) is the national norm.

Criterion-referenced tests are standardized tests that compare a student's performance to

clearly identified learning tasks or skill levels. The basis of comparison is to a body of content knowledge and skills. A CRT score displays the number of questions answered correctly and thus, presumably, how much content the test-taker has mastered.

A *pretest* is given to a child selected for Chapter 1 before Chapter 1 services begin for the program year. A *posttest* is administered after services are provided in a program year. The change between the pretest and the posttest is presumed to be an indicator of how much a child has learned during participation in the Chapter 1 program. Federal regulations require the pretest and the posttest to be administered on an annual cycle; thus, the pretest would be given in the fall (or spring) of 1991, and the posttest would be given in the fall (or spring) of 1992.

Performance assessments require students to undertake an action or create a product that demonstrates their knowledge or skills. Performance assessment can take many different forms, including writing short answers or essays, doing mathematical computations, conducting an experiment, presenting an oral argument, or assembling a portfolio of representative work. Performance assessment requires the student to produce an answer rather than simply to select one from an array of multiple choice answers.

Alternative assessment usually refers to formats other than the paper-and-pencil, multiple-choice formats used for most conventional norm-referenced or criterion-referenced tests. The term is sometimes used interchangeably with *performance assessment*. Alternative assessment may cover many different methods, including portfolios, writing samples, essay tests, computer-adaptive testing, open-ended questions or problems, presentations, or projects. The best alternative assessments engage students in tasks that provide real-world contexts or simulations, require complex and conceptual thinking, are meaningful and challenging, and evaluate both process and product.

Matrix sampling is a testing approach in which students take a sample of test items rather than an entire test; in other words, not all students are asked all the questions on the test. Scores are reported for groups rather than for individuals. This approach provides broad coverage of the subject being assessed while minimizing the testing time required of any one student.

Content standards, as used in this report, set forth the subject matter knowledge, skills, and understandings that schools should teach students to help them become well-educated and productive citizens.

Performance standards, as used in this report, establish proficiency levels or benchmarks expected of students at various ages, grades, or stages of educational development.

Delivery standards, as used in this report, indicate the conditions, resources, and inputs that contribute to a high-quality, meaningful educational experience for Chapter 1 children.

SOURCES: U.S. Office of Technology Assessment, *Testing in American Schools: Asking the Right Questions* (Washington, DC: U.S. Government Printing Office, 1992); L.M. Rudner, J.C. Conoley, and B.S. Plake, *Understanding Achievement Tests: A Guide for School Administrators* (Washington, DC: American Institutes for Research, 1989); Beth Sinclair and Babette Gutmann, *A Summary of State Chapter 1 Participation and Achievement Information: LEA Grant and State Neglected and Delinquent Programs--1989-90* (Rockville, MD: Westat, Inc., 1992); Linda Bol and Edward H. Haertel, "New Directions Toward the Development of a National Accountability Component in Chapter 1 Evaluation," paper prepared for the Advisory Committee on Testing in Chapter 1, November 1992; Ellen M. Pechman, "Use of Standardized and Alternative Tests in the States," paper prepared for the U.S. Department of Education, July 1992; Ina V.S. Mullis et. al., *The NAEP Guide* (Princeton, NJ: Educational Testing Service, 1991); and Commission on Chapter 1, *Making Schools Work for Children in Poverty* (Washington, DC: Council of Chief State School Officers, 1992).

REINFORCING THE PROMISE: THE CASE FOR REFORM OF CHAPTER 1 TESTING

Introduction

Chapter 1 of the Elementary and Secondary Education Act of 1965 helps local school districts meet the special needs of educationally disadvantaged children in low-income areas. With an appropriation of $6.1 billion for fiscal year 1993, Chapter 1 is the largest federal school aid program. Nearly all the school districts in the nation receive Chapter 1 funds, which they channel to schools with the greatest concentrations of low-income children. About half the nation's public and private schools, including 71 percent of elementary schools, participate in Chapter 1.[1] Within schools, children are selected for services in accordance with educational need, not family income. More than 5.5 million students--from a range of ethnic, racial, socioeconomic, and linguistic backgrounds--receive supplementary instruction through Chapter 1, mostly in reading and mathematics.

Since the program's early years, testing has been a routine part of Chapter 1. Chapter 1 (originally called Title I) was one of the first federal social programs to require evaluations.[2] *Standardized tests*,[3] primarily *norm-referenced tests*, have become an integral part of the program, fulfilling many functions and carrying a great deal of influence. In fact, Chapter 1 *testing procedures* affect many of the service delivery and instructional decisions in the program.

Recently new questions have arisen about whether current Chapter 1 testing procedures adequately assess the progress of children and schools, encourage the most effective instructional practices, and reflect new knowledge about teaching, learning, and *assessment*. The appointment of the Advisory Committee on Testing in Chapter 1 is one outgrowth of the interest in this issue. A consensus has emerged that Chapter 1 testing requirements need to be revised--a consensus with which this committee agrees.

To understand why Chapter 1 testing has become such a critical issue and why the committee recommends substantial revisions in Chapter 1 testing requirements, it is useful to review the challenges confronting Chapter 1 as a whole, the history and role of testing in Chapter 1, and the consequences of current Chapter 1 testing practices.

---

[1]U.S. Department of Education, *National Assessment of the Chapter 1 Program: The Interim Report* (Washington, DC: U.S. Department of Education, 1992) p. 152.

[2]Milbrey W. McLaughlin, *Evaluation and Reform: The Elementary and Secondary Education Act of 1965, Title I* (Cambridge, MA: Ballinger Publishing, 1975), pp. 1-2.

[3]Terms that are italicized the first time they appear are defined in the *Glossary of Chapter 1 Testing Terms*.

New Challenges for Chapter 1

Because of its size, reach, and longevity, Chapter 1 has exerted a major influence on American education. Chapter 1 has signaled to schools that the needs of disadvantaged children deserve sustained attention, extra funding, and special instructional efforts. It has also helped narrow the achievement gap between disadvantaged children and their more advantaged peers, at least in basic skills. Minority children in particular have made considerable gains during the period of the program's existence.[4] And Chapter 1 has promoted equity in education by redistributing resources to the nation's poorest schools.

As the reauthorization of Chapter 1 approaches, the program is facing three challenges very different from those that framed earlier reauthorizations.

First, the need for Chapter 1 is growing. About one in five school-age children now lives in poverty, 6 percent more children than a decade ago,[5] and other factors associated with educational risk, such as immigration and other health and social stresses, also are rising. These children at risk--the same children Chapter 1 seeks to help--will make up a larger share of school enrollments and, eventually, of the work force. To compete successfully for high-paying jobs, the Chapter 1 students of today and tomorrow must leave school with high-level, adaptable skills and knowledge.

The modest average improvements in basic skills that Chapter 1 children now demonstrate will not be enough to close the achievement gap. If the nation is to avoid becoming further divided into those who are educationally prepared and those who are not, Chapter 1 must be reauthorized and strengthened.

Second, education reforms are raising the achievement expectations for all children, including Chapter 1 children. The adoption of the National Education Goals and the continuing public concern about educational quality have accelerated efforts to establish high standards and a shared national vision for what children ought to know and be able to do in key subject areas. Many school districts are already voluntarily implementing the mathematics education standards developed by National Council of Teachers of

---

[4]The federally supported National Assessment of Educational Progress began measuring student achievement in reading in 1971 and in mathematics in 1973. For trend data on minority children, see Ina V.S. Mullis et al., *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading 1971 to 1990; and Writing, 1984 to 1990* (Princeton, NJ: Educational Testing Service, 1991), p. 10.

[5]U.S. Department of Commerce, *Poverty in the United States, 1991* (Washington, DC: U.S. Government Printing Office, 1992).

Mathematics.[6] Other professional organizations and expert panels, including the National Education Goals panels, are developing voluntary national standards and assessments for other subjects. The National Council on Educational Standards and Testing, a congressionally mandated panel, endorsed the voluntary adoption of national standards in five core subjects and of assessments to measure progress toward the standards.[7] Chapter 1 children need to be prepared to meet these standards.

Third, knowledge about cognition, teaching, and effective school organization has expanded dramatically in the past two decades, changing the broader educational context in which Chapter 1 operates. For example, the "constructivist" model of learning recommends that teachers build upon the understandings and conceptual experiences that children bring with them to the classroom and encourage children to connect new information with what they already know. Another body of research has enhanced awareness of what makes schools effective learning organizations.[8] Some researchers and practitioners are applying these new approaches to Chapter 1, through models that challenge disadvantaged learners with high expectations, constructivist views of learning, and opportunities for critical thinking. These models aim to demonstrate that children from all economic circumstances are capable of learning far more in addition to basic skills and that the achievement gap can be closed through early intervention.

All these challenges are causing educators and policymakers to rethink some of Chapter 1's basic provisions, including provisions that govern which schools receive funds, which children receive services, how funds are used, how fiscal integrity is maintained, and how programs are evaluated. Most of these requirements were added to the law during the 1970s, to ensure that the program served the children with the greatest needs and supplemented state and local resources. Although subsequent amendments have removed a few of the more prescriptive requirements and highlighted the flexibility available for local decision making, most Chapter 1 process requirements remain institutionalized at the state and local levels. Several recent analyses contend that together these requirements have created a Chapter 1 environment that is oriented toward regulatory compliance more than toward educational outcomes, and that encourages unproductive instructional practices, such as a remedial orientation, separate grouping of children, fragmentation of the school day,

---

[6]See National Council of Teachers of Mathematics, *Curriculum and Evaluation Standards for School Mathematics* (Reston, VA: National Council of Teachers of Mathematics, 1989).

[7]National Council on Education Standards and Testing, *Raising Standards for American Education* (Washington, DC: National Council on Education Standards and Testing, 1992), p. 2.

[8]Willis D. Hawley, "Using Chapter 1 to Fundamentally Revision Schools and Learning," paper prepared for the Advisory Committee on Testing in Chapter 1, September 1992, p. 3.

compartmentalization of school staff, and drill-and-practice instruction.[9]

Many argue that for Chapter 1 to succeed in the future, it must keep pace with reforms in education overall and, if possible, become a stimulus for further reform. Because testing requirements are inextricably tied to program operations, reform of Chapter 1 testing must be part of this process.

The Role of Testing in Chapter 1

The reliance on standardized tests for Chapter 1 has historical roots. Long before Chapter 1, education reformers saw tests as an efficient and scientific way to classify students, to measure student learning, and to hold schools accountable for results.[10] In keeping with this view, the original Title I law required school districts to use objective measures--by and large, standardized norm-referenced tests--as a means of assuring accountability for federal funds and providing federal policymakers with information about program effectiveness.[11] But because local testing practices were highly variable, and norming groups and test content were not necessarily compatible, the results of these early *evaluations* were difficult to synthesize into a national picture.

Dissatisfied with the quality of these evaluations, Congress in 1974 directed the U.S. Office of Education to develop a uniform, test-based system of evaluation and accountability--eventually called the Title I Evaluation and Reporting System (TIERS)--to meet the national, state, and local needs for information.[12] TIERS continues to govern the program today.

Revisions over Chapter 1's history have made its accountability requirements more standardized, technically sophisticated, and dependent on norm-referenced tests. The result is a current assessment paradigm that uses a single type of test for many widely different functions.

---

[9]Carolyn D. Herrington and Martin E. Orland, "Politics and Federal Aid to Urban School Systems: The Case of Chapter 1," *Politics of Education Association Yearbook 1991*, p. 177; and Linda Darling-Hammond, "Federal Policy Options for Chapter 1," unpublished paper, 1992, p. 2.

[10]U.S. Office of Technology Assessment, *Testing in American Schools, Asking the Right Questions* (Washington, DC: U.S. Government Printing Office, 1992), p. 103.

[11]Eva L. Baker and Regie Stites, "Trends in Testing in the USA," *Politics of Education Association Yearbook, 1990*, p. 145.

[12]Another name for TIERS is the Chapter 1 Evaluation and Reporting System (CHIERS).

Tests in Chapter 1 "are expected to meet needs related to teaching, program administration, and policy--and to be meaningful at the level of individual students, school buildings, and the nation as a whole."[13] As shown in the box summarizing the current uses of testing in Chapter 1, many major program decisions--including who is served, how long they are served, how much money schools receive, and which schools must modify their programs--depend somewhat or exclusively on norm-referenced *test scores*. For some of these decisions, the federal government requires the use of norm-referenced test scores, either alone or in conjunction with other data. In other cases, such as identifying eligible students, local districts use norm-referenced tests because of long-standing Chapter 1 practice or local convenience.

---

[13]Brenda J. Turnbull, "Testing in Chapter 1: Issues and Options," paper prepared for the National Assessment of Chapter 1 Independent Review Panel, July 1991, p. 1.

CURRENT USES OF TESTING IN CHAPTER 1

- Student diagnosis. Teachers, at their discretion, may use tests and other assessments to diagnose student learning needs.

- Instructional feedback. Teachers may use tests, along with other formal and informal assessments, to gauge student progress and to judge the effectiveness of instruction.

- Student eligibility and identification. School districts must use educationally related objective criteria, including test scores, to identify which children are eligible for Chapter 1 and to select those with the greatest needs for services.

- Needs assessment. School districts must determine which grade levels and subject areas to cover in Chapter 1 by annually assessing the needs of eligible children. Student scores on standardized tests are a major source of these data.

- Funds allocation. School districts must use educational need criteria--which often mean test data--to determine how much Chapter 1 money each school should receive.

- Local accountability. To fulfill federal evaluation requirements, school districts must annually test Chapter 1 children in basic and more advanced skills and must submit aggregated scores to the state.

- Student program improvement. School districts must use test scores, alone or with other measures, to determine which children have been in Chapter 1 for two years without making sufficient progress, to modify instruction for these children, and to decide when children are no longer eligible for Chapter 1.

- School program improvement. School districts must use test scores to determine which schools need to engage in the program improvement process and how long the school will stay in it.

- Schoolwide project continuation. School districts must use test scores to determine whether schoolwide projects can operate beyond three years.

- State accountability. States must aggregate local test data and submit it to the U.S. Department of Education to meet federal evaluation requirements.

- National accountability. Congress and federal program administrators use aggregated test data to evaluate Chapter 1's effectiveness, to weigh policy changes, justify funding levels, and to target states and districts for federal recognition and monitoring.

Through laws, regulations, and guidelines, the federal government not only mandates norm-referenced tests in Chapter 1, but also prescribes national models and procedures, through TIERS, for how and when to administer tests and how to interpret and report results. The purpose of these requirements is to produce local test scores that are reasonably comparable and uniform, so that they can be aggregated into a national portrait of Chapter 1 student achievement. TIERS currently includes the following requirements:

- School districts must assess the achievement of Chapter 1 students in grades 2 through 12 on an annual cycle (fall-to-fall or spring-to-spring), using a *nationally normed test* or a state or local test that can be equated to national norms.

- School districts must calculate student achievement gains in basic and "more advanced" skills and express these gains in *normal curve equivalents* (NCEs), a statistic that resembles a percentile rank but permits aggregation of results. NCE gains occur when Chapter 1 children improve their rank relative to the test's norm group.[14] The requirement to assess more advanced skills (higher-order skills) has been met by administering certain subtests of standardized tests.

- At least once every three years, school districts must determine whether Chapter 1 student achievement gains are sustained for a period of more than one program year by testing the same children for at least two consecutive 12-month periods.

- At least once every three years, school districts must also assess the progress of Chapter 1 children in the regular instructional program by reviewing test scores or other relevant indicators.

- School districts must report aggregated Chapter 1 achievement scores to the state education agency (SEA), which in turn reports them to the U.S. Department of Education.

Nearly all Chapter 1 children in grades 2 through 12 take a norm-referenced test once a year to fulfill federal evaluation requirements. In program year 1989-90, at least 1,642,000 Chapter 1 students had taken both a *pretest* and *posttest* in reading and at least 995,000 students had taken pre- and posttests in mathematics.[15] Because these

---

[14]The term NCE "gain," which is commonly used in Chapter 1, signifies a change in the relative rank of a Chapter 1 student and may or may not signify a gain in absolute achievement.

[15]There is some overlap between the reading and mathematics students. Beth Sinclair and Babette Gutmann, *A Summary of State Chapter 1 Participation and Achievement Information: LEA Grant and Neglected and Delinquent Programs--1989-90* (Rockville, MD: Westat, Inc., 1992), p. 47.

represent only the students for whom districts reported matched pre- and posttest scores a year apart, the actual number tested is much higher.

The data generated through this testing form the basis of current local, state, and federal Chapter 1 evaluations and are used for other decisions, such as identifying schools for program improvement and judging schoolwide projects. But it is the information needs at the federal level that drive the reliance on norm-referenced tests as the primary measure of achievement.

Standardized norm-referenced tests have several advantages as a basis for national accountability in Chapter 1. They produce reliable, valid information on many types of tasks, diminish problems of teacher bias, are relatively inexpensive, are comparatively easy to administer, can be scored by machine, and are developed in accordance with high technical standards. In addition, they provide an external check on local curriculum and instruction and a "remote control" that policymakers, parents, and citizens can use to monitor systems or leverage change.[16] In the aggregate, standardized test scores can also be a readily grasped symbol of school success or failure: "A steep trend line on a graph can be strong ammunition in political struggles over the quality of schools."[17] Norm-referenced tests have the added advantage of producing scores that can be aggregated.

Reexamining Current Testing Procedures

Several developments have engendered new questions about the current Chapter 1 testing procedures, especially their reliance on a single measure for accountability.

First, a revolution is taking place in the testing and measurement field, as researchers, testing experts, and educators develop and refine new assessment formats that are better aligned with state and local curricula, give students tasks with real-world value, and yield information about a test taker's reasoning processes. For instance, 34 states have implemented or are piloting *performance assessment* or *alternative assessment* formats--most often, direct writing assessments, but also essay tests, tests with open-ended questions, scientific experiments, exhibitions, portfolios of student work, and computer- and video-administered testing.[18] Although these state assessments are at various stages of readiness, they signal a change of thinking and a desire for approaches different from those currently available.

---

[16]Larry Cuban, "The Misuse of Tests in Education," paper prepared for the U.S. Office of Technology Assessment, OTA contractor report PB 127653, September 1991, p. 9.

[17]U.S. Office of Technology Assessment, *Testing in American Schools*, p. 54.

[18]Ellen M. Pechman, "Use of Standardized and Alternative Tests in the States," paper prepared for the U.S. Department of Education, July 1992, p. 9.

Second, there is a growing awareness that Chapter 1 testing requirements influence not only what happens in Chapter 1, but also how much and what kind of overall testing occurs in schools. Because states and local districts often use the same tests to meet both Chapter 1's needs and their own needs, and because state and local testing programs have expanded during roughly the same period as Chapter 1, it is difficult to isolate the effects of Chapter 1 on local testing, let alone compute the additional costs. Even so, many school districts report that they would do less testing or different types of testing without Chapter 1.[19] Many school districts, especially large urban ones, find it easier to test all students on a schedule that meets Chapter 1 demands than to cull Chapter 1 students for separate testing. When state and locally driven testing does not cover every grade, Chapter 1 may require a duplicate testing program.[20] Although federal regulations technically permit school districts to use state or local tests that can be equated to national norms, this type of waiver has been requested and granted in only a few cases.

Third, the 1988 Hawkins-Stafford School Improvement Amendments (P.L. 100-297) raised the stakes and multiplied the uses of testing in Chapter 1. New program improvement provisions required schools with stagnating or declining average test scores among Chapter 1 students to modify their programs. A revamped schoolwide projects option made it easier for schools with high poverty to use Chapter 1 funds to upgrade instruction for the entire student body, but only if these schools could demonstrate improved Chapter 1 test scores after three years. Other provisions established new goals for Chapter 1--to help disadvantaged children perform at grade level, develop higher-order thinking skills, and succeed in the regular instructional program. The overall intent of the amendments was to focus more attention on program quality and student outcomes. The amendments did not, however, change the methods used to measure these outcomes in any substantial way; as a result, the limitations of the current system have become more obvious.

The Consequences of Chapter 1 Testing Procedures

The committee has analyzed the effects of the current requirements and concludes that the reliance on standardized, norm-referenced achievement tests for many Chapter 1 assessment functions, in combination with other Chapter 1 requirements, produces several dysfunctional consequences that point to a need for major revisions. These dysfunctional consequences can be summarized as follows.

---

[19]U.S. Office of Technology Assessment, *Testing in American Schools*, pp. 85-86.

[20]Lorrie A. Shepard, "Chapter 1's Part in the Juggernaut of Standardized Testing." Paper prepared for the annual meeting of the American Educational Research Association, April 1992, p. 3.

Failure to Employ Multiple Measures

The practice of relying on the results of a single test or assessment device to make important decisions about instruction or students is at odds with a widely accepted principle of educational measurement: "No one test can do it all. The multiple measurement approach to assessment is the keystone to valid, reliable, and fair information."[21]

In the area of student identification, for example, valid norm-referenced tests can provide useful information for identifying the lowest-achieving children from among a larger group. When used in conjunction with other indicators, as the current law and regulations permit, they can aid school districts in making sound judgments about which children are in greatest need of Chapter 1 assistance. Some schools, however, use norm-referenced tests exclusively to select Chapter 1 children for participation, without bringing to bear other indicators.[22] Yet the educational needs of children can be quite complex. Additional measures, such as the informed professional judgment of teachers, can provide valuable information about the needs of children for Chapter 1 services, which could help interpret or enrich the information available from conventional tests.

The 1988 amendments served to highlight the difficulties inherent in using a single measure to judge the effectiveness of Chapter 1 at the school level. However mild the sanctions attached to program improvement, the use of student achievement gains on norm-referenced tests for these decisions has generated new concerns about the accuracy of school-level performance data, which may be undermining enthusiasm for the entire program improvement process.[23] Although the Department of Education has encouraged states to establish high standards to trigger program improvement, fewer than half the states have set standards above the federal minimum (NCE gains greater than 0).[24] States may be reluctant to identify more schools than they can serve with program improvement funding, and some seem to be cautious because they lack confidence in a process so dependent on norm-referenced tests.[25]

Studies point to the following weaknesses of using average gain scores on a single measure at the school level:

---

[21]Michael H. Kean, "ESEA Chapter 1 Reauthorization," presentation prepared for the Advisory Committee on Testing in Chapter 1, August 1992, p. 3.

[22]National Association of Early Childhood Specialists in State Departments of Education, cited in Turnbull, "Testing in Chapter 1," p. 20.

[23]U.S. Department of Education, *National Assessment*, p. 52.

[24]NCE = 0 means that the student did not lose ground in relative rank.

[25]U.S. Department of Education, *National Assessment*, p. 54.

- Year-to-year changes in aggregate test scores may not always be reliable indicators of program quality. Half the schools identified for program improvement "test out" of the process before implementing their plans, by scoring at sufficiently high levels during the year following their identification.[26] Sometimes these changes seem to be attributable to measurement phenomena rather than to genuine improvements in achievement.[27]

- Using a different achievement measure may produce a decidedly different picture of program effectiveness. The results of state-developed *criterion-referenced tests* more closely aligned with state curriculum sometimes contradict performance results from norm-referenced tests, yet it is the latter measure that governs program improvement and other accountability decisions.[28]

- Basing local accountability on achievement gains relative to a norm precludes consideration of the absolute level of performance of Chapter 1 students.[29]

The 1988 Amendments recognized the desirability of multiple measures for program improvement decisions by directing school districts to establish "desired outcomes" for Chapter 1, which could be assessed with measures other than norm-referenced tests. But federal regulations require schools to demonstrate both gains on the norm-referenced tests and progress toward attaining desired outcomes to avoid being subject to the law's program improvement requirements. Thus, desired outcomes become an additional hurdle instead of an alternative or complement to gains on norm-referenced tests, giving districts little incentive to choose challenging desired outcomes.[30]

Inappropriate Uses

Another basic principle of testing is this: "Tests of any kind should be used only to serve the functions for which they were designed and validated."[31] Because norm-

---

[26]Ibid., p. 8.

[27]Judith I. Anderson, "Using the Norm-Referenced Model to Evaluate Chapter 1," paper presented at the annual meeting of the American Educational Research Association, April 5, 1991, p. 11.

[28]William L. Padia, "Chapter 1 Assessment Issues," presentation to the Advisory Committee on Testing in Chapter 1, August 1992, n.p.

[29]U.S. Department of Education, *National Assessment*, p. 54.

[30]Turnbull, "Testing in Chapter 1," p. 22.

[31]U.S. Office of Technology Assessment, *Testing in American Schools*, p. 7.

referenced tests are required to fill certain assessment roles, many Chapter 1 programs are inclined, for reasons of efficiency, to use these tests to fill other roles for which they were never intended. Although norm-referenced tests certainly have an important place in the continuum of assessment, like any other measure they fulfill some purposes better than others. As one researcher concluded, "The harm is not so much in these tests themselves; the harm is in using them excessively and in situations where something else is needed."[32]

For example, while norm-referenced tests can provide some useful feedback to teachers when they are administered properly and reported in a timely and detailed way, teachers need a wide range of assessment tools to diagnose student needs, monitor children's day-to-day progress, and make decisions about altering instructional practices.[33] Of course, schools and teachers do administer other forms of assessment to Chapter 1 children besides those required by the federal government for formal evaluation. The problem is that the scores on the norm-referenced tests are the ones that count for student and school program improvement and for mandated TIERS evaluations.

Impediments to Appropriate Curriculum and Instruction

The reliance of Chapter 1 testing on currently available norm-referenced tests fails to acknowledge reforms in testing, cognitive psychology, curriculum, and instruction, and may be standing in the way of needed innovations and improvements in Chapter 1. Concern has been expressed about how well current testing procedures measure higher-order skills, as required by the 1988 amendments. Critics contend that conventional multiple-choice formats do not reveal children's thinking processes or problem-solving strategies. Some say that even those standardized tests that purport to test higher-order skills may measure only a few such abilities.[34]

Because norm-referenced tests are designed to be relatively independent of particular curricula or instructional practices, their content is unlikely to match closely state and local academic goals.[35] Consequently, the instructional feedback they yield may seem somewhat remote and of limited usefulness. Chapter 1 projects that use an early childhood developmental approach, manipulative-based mathematics instruction, active science learning, or a whole language approach may find the mismatch between instruction and testing

---

[32]David Sweet, "Reconsidering Current Federal Policy for Evaluating ESEA Chapter 1 Grants to Local Educational Agencies," paper presented to the American Educational Research Association, April 1991, p. 3.

[33]Judy Pfannenstiel, "Measures of Achievement in Chapter 1," paper prepared for the Advisory Committee on Testing in Chapter 1, September 1992, pp. 8-10.

[34]U.S. Department of Education, *National Assessment*, p. 133.

[35]Ibid., p. 77.

frustrating if student progress does not show up as gains on the test. There is incentive, in this situation, to change instruction to match the test.

There is evidence that Chapter 1 testing procedures may indeed be promoting undesirable instructional practices, limiting the kinds of learning experiences to which students are exposed, or reinforcing outmoded ways of teaching disadvantaged children. Some researchers and practitioners contend that Chapter 1 testing requirements may be narrowing Chapter 1 curriculum and instruction by rewarding those practices most likely to produce gains on norm-referenced tests.[36] There is also some evidence that an emphasis on multiple-choice norm-referenced tests may encourage teachers to spend undue time teaching test-taking skills or low-level basic skills, rather than more challenging content.[37] For example, teachers may give classroom tests or worksheets with the same format as standardized, multiple-choice tests, or may drill students in recognizing the main idea of a short reading passage rather than connecting their understanding of the story with their own experience.[38] Some analysts also contend that an overemphasis on standardized norm-referenced test gains may displace instruction in subjects not normally tested, such as science or geography.[39]

Many of the instructional practices rewarded by the use of currently available norm-referenced tests are those that Chapter 1 programs have been heavily criticized for employing. Many Chapter 1 schools remain "bastions of conventional practices that emphasize the sequential mastery of discrete basic skills."[40] In the current climate of educational reform, Chapter 1 does little service to disadvantaged children by holding them to lower expectations or giving them less challenging content than other children.

Failure to Assess Special Populations Appropriately

The current Chapter 1 reliance on norm-referenced tests does not accommodate the special assessment needs of particular groups of students. One such group consists of young children, defined by the committee as those below grade 3. For example, many early childhood education professionals and researchers conclude that young children often perform poorly on group-administered, paper-and-pencil tests in contrived settings because these children have limited attention spans and manual dexterity. Moreover, critics contend that the development of children is highly episodic; a test given one day may not reflect what a

---

[36]Turnbull, "Testing in Chapter 1," p. 12.

[37]Baker and Stites, "Trends in Testing in the USA," p. 151.

[38]Shepard, "Chapter 1's Part in the Juggernaut of Standardized Testing," p. 2.

[39]Turnbull, "Testing in Chapter 1," p. 12.

[40]U.S. Department of Education, *National Assessment*, p. 33.

child will know or be able to do just days later. Furthermore, some argue, the tests often used to assess young children have methodological weaknesses.[41]

Although Chapter 1 regulations do not require the use of standardized testing below grade 2, some districts administer such tests at these grade levels to determine student eligibility or for other Chapter 1 purposes. Available norm-referenced tests for young children do not fare well when judged from the currently accepted model of developmentally appropriate education for young children. Developmentally appropriate assessment should occur in a variety of settings over a period of time, should closely resemble classroom activities, and should be based primarily on observational data from teachers, parents and performance samples.[42]

Limited-English proficient (LEP) students and children with learning disabilities constitute two other groups with special assessment needs. Many LEP students do not possess sufficient English language skills to permit their being appropriately assessed with the written testing instruments commonly used in Chapter 1. Moreover, oral language assessment tools are not sanctioned in Chapter 1 testing. Chapter 1 regulations allow school districts to identify and serve LEP children or children with disabilities if the children (1) have needs stemming from educational deprivation and not related solely to their limited English proficiency or handicapping condition, and (2) are selected on the same basis as other children. It is common, however, to find LEP and special education students exempted from testing, and this situation presents a dilemma. On the one hand, children who are not assessed can easily go unnoticed and receive no services. On the other hand, it makes little sense to assess a child if he or she lacks sufficient language proficiency to take the test being used.

Questionable Validity and Utility of Results

The current Chapter 1 testing procedures were designed, first and foremost, to yield useful national data on which to base policy, funding, and program decisions. Yet evidence suggests that even this goal is not being met effectively.

From a technical perspective, collecting test scores on every student receiving Chapter 1 services and aggregating them upward is a singularly inefficient way of gauging

---

[41]Samuel J. Meisels et al., "Testing, Tracking, and Retaining Young Children: An Analysis of Research and Social Policy," commissioned paper, 1989.

[42]National Association for the Education of Young Children, "Position Statement: Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 through 8," *Young Children*, March 1991, pp. 32-33.

overall program success.[43] TIERS depends on the testing efforts of thousands of school districts with varying expertise in measurement—a situation that risks awesome quality control problems. State and local performance reports submitted for Chapter 1 evaluation sometimes contain implausible or incomplete data, and local coordinators sometimes seem unaware of the proper procedures for compiling pretest and posttest data.[44]

In addition, TIERS falls well short of obtaining data for all Chapter 1 participants. In 1989-90, states reported "matched" pretest and posttest scores for only 60 percent of Chapter 1 reading participants and 55 percent of Chapter 1 mathematics students.[45] Student mobility is the major reason school districts cite for the absence of matched test scores. Because the more mobile students are likely to differ in significant ways from other Chapter 1 students, TIERS data may not be representative.

According to a recent study, many states treat the collection of performance data as a "routine, bureaucratic" ritual, seldom subject to independent analysis.[46] States do not seem to have complete confidence in or make extensive use of their own TIERS data.[47] Because of incompatible federal and state requirements, Chapter 1 is missing an opportunity to link up with the reforms that many states are pioneering in the areas of assessment and curriculum.

Congress itself does not appear to be thoroughly satisfied with the TIERS data.[48] To prepare for reauthorization, Congress usually mandates several special studies on the effects of Chapter 1 that often are more widely cited than the TIERS reports.

---

[43]R.E. Barnes and A.L. Ginsburg, "Relevance of the RMC Models for Title I Policy Concerns," *Educational Evaluation and Policy Analysis* 1 (March-April), quoted in U.S. Department of Education, *National Assessment*, p. 84.

[44]U.S. Department of Education, *National Assessment*, pp. 82-83.

[45]Six states are not included in this data. Sinclair and Gutmann, *A Summary of State Chapter 1 Participation and Achievement Information*, p. D-58.

[46]U.S. Department of Education, *National Assessment*, p. 74.

[47]Ibid., p. 9.

[48]Ibid., p. 84.

REFORMING THE PARADIGM: RECOMMENDATIONS FOR CHAPTER 1 TESTING

The Purpose of the Advisory Committee

The aforementioned concerns about testing led the U.S. Department of Education to establish the Advisory Committee on Testing in Chapter 1. The committee is one of several national panels studying Chapter 1 issues to prepare for the upcoming 1993 reauthorization of the law. It is the only study group solely devoted to examining testing requirements and, as such, is uniquely positioned to provide analyses and alternatives that the Congress and the Secretary can use to weigh policy options.

The committee's charter (see Appendix A) directs it to undertake the following tasks:

- help the Secretary assess the adequacy of standardized tests used in Chapter 1;
- conduct a comprehensive examination of the standardized tests currently used;
- advise the Secretary on possible improvements or alternatives for current testing procedures, including possible changes in test instrumentation, administration, and the reporting of test results at the state and local levels;
- provide guidance on national standards that may permit aggregation of data and measurement of program impact without depending exclusively on norm-referenced testing; and
- make recommendations to the Secretary on regulatory or legislative changes that will allow an alternative assessment procedure and provide information upon which schools, districts, and states can be held accountable for student performance.

To fulfill these charges, the committee held eight days of meetings and commissioned several policy papers (see Appendix B).

After analyzing the effects of current Chapter 1 testing procedures and debating alternative approaches, the committee is proposing a new paradigm for Chapter 1 testing and assessment, based on a set of guiding principles and implemented through a set of specific recommendations.

Guiding Principles

The committee has arrived at several principles that should guide reform of Chapter 1 testing at all levels of government and for all purposes. These principles form a philosophical basis for the specific recommendations in the next section.

Principle #1: Chapter 1 should continue to have strong accountability, but the balance should shift to emphasize how well students are learning and how effectively they are being taught.

Over the next five years, Chapter 1 should make a strategic transition to an outcome orientation in assessment and in this way become a lever for reform in education. The paradigm we propose would recalibrate the testing balance so that the instructional information needs of teachers and the learning needs of students receive at least as much weight as the compliance and accountability needs of administrators and policymakers.

The committee is not proposing a reduction in accountability or a dilution of those features that have successfully targeted Chapter 1 on the neediest children. Chapter 1 has evolved to the point, however, that state and local administrators understand its purpose and respect its integrity. The committee's recommendation would allow states and local districts a new degree of flexibility and creativity, but in exchange the states and districts would have to demonstrate that Chapter 1 children were improving their achievement and that schools were providing effective programs. Classrooms, schools, and systems would still be held accountable for educational progress, but the measures by which progress is judged would be more closely linked to the outcomes sought.

Because of its size and importance, Chapter 1 has the potential to leverage reform in all of education. In the paradigm we propose, Chapter 1 would encourage development of high expectations for all children, stimulate research and development in the assessment field, demonstrate ways that disadvantaged children can succeed in the regular school programs, and help Chapter 1 schools become model learning organizations that will better serve all students.

All this will take time, however. A transition period of up to five years would be provided (see Recommendation #7 below), along with technical and financial assistance, for Chapter 1 to move to a new outcome-based orientation.

Principle #2: Chapter 1 testing should no longer be an independent system but should be linked with the education reforms that states and school districts are undertaking for all *children.*

Because the ultimate goal of Chapter 1 is to help disadvantaged children succeed in the regular school program, Chapter 1 children are not well served by being held to lower expectations or being given less challenging content than other children. Spurred by the National Education Goals and the efforts of several professional associations and study groups, the nation is already debating and developing voluntary national standards for what all children should know and be able to do in critical subjects. Within the next few years, national standards will probably be developed for some or all of the subjects mentioned in the National Education Goals and ready for voluntary adoption. Whether states and local districts adopt these standards or establish other expectations for student achievement, the

performance outcomes for Chapter 1 children should be the same as those for all children. Toward this end, Chapter 1 testing should be linked--from the classroom to the state level--with the curricular objectives, performance standards, and assessment criteria applicable to all children. Chapter 1 assessment should also encourage schools to integrate Chapter 1 services as fully as possible into the regular instructional program.

Principle #3: National Chapter 1 evaluation should be decoupled from state, local, and classroom assessment functions.

As an organizing principle for its recommendations, the committee has grouped the functions of Chapter 1 testing into five major categories: (1) national accountability, (2) state accountability, (3) local accountability, (4) instructional feedback and diagnosis, and (5) student eligibility and identification.[49] We propose that each of these functions be met with a different combination of appraisal procedures.

Under the committee's paradigm, Chapter 1 testing would become primarily a state and local activity, with federal guidance. National accountability needs would be met with an independent national assessment, giving Congress more latitude to design a national evaluation that addressed its specific needs and to revise other components of Chapter 1, such as program improvement, that are currently constrained by national testing procedures. Freed from the demands of national aggregation, states, school districts, and teachers would have greater flexibility to develop combinations of assessments that best suited their particular needs, within federal guidelines that ensured sound practice but allowed for variations. The incentives would be reduced for states and local districts to use tests for purposes for which they were not designed or to use assessments ill matched to their curricular and instructional goals. The result should be greater flexibility and better assessments at all levels.

The combinations of assessments for each of the five functions of testing would be distinct but interrelated. For example, teachers might receive instructional feedback from the assessments used for state and local accountability and from other assessments they selected themselves. As national standards become available in the various subjects, all levels of assessment should eventually be related.

To the extent possible, the recommended Chapter 1 assessments should be developed in a way that lessens the burden and cost of testing. If implemented as envisioned, the committee's proposals could result in less testing than now takes place. National evaluations would be done on a sampling basis. States and local districts could use sampling procedures

---

[49]In this report, "student eligibility" refers to the process used to determine which students qualify for Chapter 1 services; "student identification," sometimes called student selection, refers to the process used to select which students will receive Chapter 1 services when there are insufficient funds to serve all who are eligible.

if they were able to develop a valid, representative sample. They could also test less than annually, as long as the assessment schedule provided sufficient information to ensure accountability for students and schools.

Principle #4: The diverse purposes of assessment in Chapter 1 should be met with multiple methods of assessment.

Chapter 1 should require multiple assessment methods at all levels of government and for all major program decisions. These should include alternative assessments where appropriate.

Some states have progressed considerably in developing alternative forms of assessment; other states have not. And although there are many unresolved issues related to alternative assessments--among them ethnic and gender bias, cost and reliability of scoring, and level-to-level scale linkages--alternative forms of assessment have the potential to tap important elements of student learning that conventional multiple choice tests cannot.

The benefits and limitations of all major forms of assessment--performance assessment, developmentally appropriate assessments, computer- and video-adaptive assessments, essay and constructed-response tests, norm-referenced tests, and criterion-referenced tests--should be weighed. In designing the right mix of assessments, federal, state, and local agencies should take into account the purposes for which various assessment formats have been designed and validated, their appropriateness or inappropriateness for high-stakes decisions, their match with state and local curricular and instructional goals, and their capabilities to assess the various outcomes expected for all children.

By encouraging alternative assessments, the committee hopes to build on the testing and measurement reforms already occurring at the state and local levels and in research laboratories, and in this way benefit all students. We do not intend to promote the use of alternative assessment for decisions for which they are inappropriate, or in cases where further research and refinement is needed.

Norm-referenced tests and criterion-referenced tests could still have a place in the new paradigm. If state and local education agencies (LEAs) determine that norm-referenced tests fulfill some important needs, they should continue using them, with appropriate controls to prevent misuse. In addition, SEAs and LEAs should consider alternative assessments to evaluate skills and knowledge--such as writing, oral language (listening and speaking), and strategies for solving complex problems--that conventional tests do not assess as well.

Recognizing the steps that still must be taken before some forms of alternative assessment are ready for wide implementation, the committee recommends that the federal government support staff development and research and development on assessment (see Recommendation #8 below).

Principle #5: Chapter 1 assessment should recognize the different developmental stages of children.

The committee supports early intervention in Chapter 1. It is absolutely critical to intervene at the preschool and early primary years so that disadvantaged children receive a solid educational grounding. Care must be taken, however, to assess young children in appropriate ways.

Recognizing the cautions raised by early childhood experts regarding the testing of young children, the committee proposes different types of assessment strategies for children and programs at the following age and grade levels: prekindergarten and kindergarten; grades 1 and 2; and grades 3 through 12. The recommendations in the next section clarify the assessment, program delivery, and identification and selection procedures for young children. A basic principle is that Chapter 1 should use alternative assessments that are developmentally appropriate for children below grade 3,[50] and that Chapter 1 accountability for prekindergarten and kindergarten programs should be based only on how well the program is being delivered.

Recommendations

To implement the new paradigm, the committee offers eight recommendations for Chapter 1 assessment. Recommendations #1 through #5 are organized around the five major functions of Chapter 1 testing. Recommendations #6 through #8 cut across several functions and are necessary steps to ensure the success of the other recommendations.

National Accountability

Recommendation #1. The federal government should periodically evaluate the national effectiveness of Chapter 1 using a NAEP-like assessment that evaluates the achievement of a representative sample of Chapter 1-eligible students beginning in grade 3; collects background information on Chapter 1 students and programs; and analyzes the long-term effects of Chapter 1 participation. Special procedures should be used to ensure national accountability below grade 3.

The committee recommends that the federal government begin immediately to design and implement a national assessment for Chapter 1 that will meet the accountability demands of federal policymakers and provide national information about the program's effectiveness. The assessment should be designed, operated, and administered using the same scientific principles, technical procedures, and quality control standards that govern the National

---

[50]Current Chapter 1 testing practices do not require norm-referenced testing for children below grade 2.

Assessment of Educational Progress (NAEP). Through a NAEP-like *matrix sampling* approach, the assessment could yield more accurate achievement and background information at the national level than the current TIERS approach of national aggregation, while testing a fraction of the number of students and perhaps costing less.

The national assessment should appraise the performance of a representative number of participating and nonparticipating Chapter 1-eligible children in Chapter 1 schools, and Chapter 1-eligible children in non-Chapter 1 schools. A representative sample of schoolwide projects should be included. The sample should be sufficiently large to generalize to important linguistic (including LEP), ethnic, and other subpopulations of students. The committee recommends that special procedures be undertaken to ensure that LEP students are not excluded from the sample.

The national Chapter 1 assessment should be conducted on a multi-year cycle rather than annually; every two to four years is suggested. Rather than assessing students in all the grades 2 through 12, the national assessment should periodically target specific grades, beginning in grade 3. For example, the assessment might be conducted at grades 3, 5, and 8--enough to cover the span of Chapter 1 but not so much as to be unnecessarily costly and burdensome.

The assessment should measure the performance of Chapter 1 children in reading, writing, oral language, mathematics, geography, history, and science (although not every subject would have to be assessed every time). Student performance should be assessed in terms of proficiency levels, rather than against a norm. It would also be useful if the Chapter 1 assessment produced scores similar to those produced by NAEP, so that comparisons could be made to performance levels of the broader student population over time. The proficiency levels for the assessment should reflect voluntary national standards for the subjects tested as they become available; until national standards are available, the assessment could use NAEP items, content, and proficiency levels.

The first assessment done for national accountability should be conducted on a state-by-state basis, to establish a baseline. The committee recognizes that state-by-state samples are more expensive than a national sample; to control costs, subsequent assessments would not always have to be conducted state by state. The committee suggests that a state-by-state assessment be conducted every four to six years, in accordance with the congressional Chapter 1 reauthorization schedule, so that the states and the Congress will have state-level data to supplement the information emerging from the state-developed assessments proposed below in Recommendation #2. If, in the years that the national sample is not conducted state by state, the sample still yields sufficient data to produce valid results for state-level analysis in the largest states, then these data should be made available to those large states that desire the data.

Because the information from this national assessment is not meant to be used below the state level, the committee recommends that no data be analyzed or reported at the school,

classroom, or student level. As explained later, other assessment methods more closely tied to state and local curricula would be used to fulfill Chapter 1 state, local, and classroom assessment needs.

The committee recognizes the need for some form of national accountability for Chapter 1 children from prekindergarten through grade 2, as well as the sensitivities about testing young children. Therefore, the committee recommends that the Secretary of Education systematically review data from Chapter 1 programs at the prekindergarten through grade 2 levels in every state, to determine how effectively the program is being delivered and to ensure that states are not neglecting the interests of these children.

Standardized performance data would not be required for national accountability at the preschool, kindergarten, or first grade levels. However, standardized performance data may be collected in grade 2 if the following criteria are met:

- assessments are performance-based, not paper-and-pencil tests, conducted in the child's natural environment.
- data are collected at multiple points in time.
- matrix sampling is used.
- data are not used to stigmatize, label, or place any child.
- data are collected in multiple domains, including social-emotional development, language (including oral language) and literacy, and approaches toward learning.

Given the existing state of assessment for young children, the ongoing work of National Education Goals panel #1 on school readiness, and the deliberations that are necessary to accomplish the condition listed above, a panel on early childhood assessment should be established to oversee the development of such an assessment. Care should be taken to see that such a panel has the technical and practical expertise to execute the task, and that its composition is culturally diverse and responsive to the needs of language-minority and special education children.

It is further recommended that Chapter 1 early childhood programs be linked in design and databases with other federal preschool programs.

To enrich the nation's knowledge about Chapter 1 students and programs and to help interpret national achievement data, the national Chapter 1 assessment should also collect other background information about Chapter 1 at all levels, such as numbers and characteristics of children (i.e., race, ethnicity, and socioeconomic background), types of communities and schools, numbers and characteristics of teachers and specialists, staff development opportunities, per-pupil expenditures, types and intensity of services,

instructional practices, and other program indicators.

The national assessment should contain a longitudinal component that tracks a cohort of Chapter 1 students at least into young adulthood and examines the long-term effects of the program, similar to the High School and Beyond project of the National Center for Education Statistics. In addition to academic achievement, this longitudinal component ought to appraise factors such as participation in higher education, employment and earnings, dropout rates, and delinquency. To the extent possible, the national assessment should be linked to the information resulting from the congressionally mandated Prospects longitudinal study of Chapter 1 participants.

The Congress and the Department of Education should have the authority to continue mandating studies addressing special Chapter 1 issues other than achievement. In addition, the committee recommends that the federal government systematically undertake special national studies of the English proficiency (reading, writing, listening, and speaking) of LEP children in Chapter 1. All the additional studies should be coordinated with the national accountability assessment.

State Accountability

Recommendation #2. States should develop and implement: (a) content and performance standards that will address what Chapter 1 students should know and be able to do, which should be the same as standards for all students; (b) Chapter 1 assessment methods aligned with these content and performance standards; (c) delivery standards for effective Chapter 1 programs; and (d) procedures for local reporting of student assessment results and state monitoring of Chapter 1 program delivery and effectiveness.

A major aim of the committee's recommendations is to vest greater responsibility, leadership, and action for Chapter 1 assessment and accountability in state educational agencies. Because the main playing field for educational reform has been at the state level, reforms in Chapter 1 testing will be more meaningful and successful if they are linked to the improvements in curriculum and assessment that the states are already striving to make for all children. Where states have already developed outcome measures and related curriculum frameworks and assessments for all students, the committee wishes to encourage these activities. Where states have not undertaken these strategies, we recommend that Chapter 1 help launch states in this direction.

The committee specifically recommends that over a five-year transition period (see Recommendation #7 below), SEAs develop and put in the place the following: *content standards*, *performance standards*, related assessments, *delivery standards*, and reporting and monitoring procedures. Special procedures would govern program delivery and assessment for prekindergarten and kindergarten programs and programs in grades 1 and 2. The SEA would submit plans containing all standards, assessments, and procedures to the U.S.

Secretary of Education for approval.

The purpose of these standards, assessments, and procedures would be to guide local Chapter 1 assessment within the state. States could require uniform assessment procedures of all LEAs, or, at the state's option, could allow local districts to modify state models or develop their own standards and assessments subject to state review and approval. In the latter case, local agencies should still be subject to state monitoring and reporting procedures.

These standards, assessments, and procedures should have the following characteristics:

Content standards--the expectations for subject matter knowledge, skills, and understandings that schools should teach Chapter 1 students to help them become well-educated and productive citizens--should be the same as the content standards expected for all children and should include both basic and higher-order skills. The committee suggests, at a minimum, that the standards address reading, writing, oral language, and mathematics, and further assumes that states will also address science, history, and geography, because Chapter 1 children should become proficient in all the subjects addressed in the National Education Goals. To the extent possible, the standards should be compatible with whatever voluntary national standards are available in the subjects addressed. It is not the intent of the committee to undermine the voluntary nature of any national standards, however.

States could also elect to cover subjects not addressed in the National Education Goals or could develop standards for interdisciplinary approaches. Furthermore, content standards should be compatible with state curricular frameworks and accreditation standards.

Performance standards--the proficiency levels or benchmarks students should attain at various ages, grades, or stages of educational development--should be ambitious; in fact, they should be the same as the performance expectations for all children. An example of a performance standard is this: By the end of fourth grade, Chapter 1 students should be able to apply estimation in working with quantities, measurement, computation, and problem solving. The specific ages, grades, or points in time would be determined by states, but the standards should adequately represent the grade levels served by Chapter 1 programs. State performance standards for Chapter 1 might also address important non-academic goals.

Related assessments--the methods by which states and local districts would measure the progress of Chapter 1 children toward content and performance standards--should form the core of Chapter 1 accountability at the state and local level, and should provide important instructional feedback to teachers (see Recommendation #4 below). Multiple measures, including alternative assessment, should be required.

State assessments for Chapter 1 should be consistent with and valid for the content and performance standards outlined earlier and similar to state assessments for all children. Assessments should also be compatible with any assessments states use to measure progress

toward voluntary national standards and with statewide assessment practices for all students. If states want to use existing assessments that are fair, sound, and in tune with effective instructional practices, they should be able to do so, as long as they are valid for the standards approved for the state. Whatever assessment instruments a state chooses or develops, the state should submit validity data and other supporting information to the Secretary of Education.

Sampling should be permitted, as long as the state can structure a sample representative of Chapter 1 children in the state. The sample should also include sufficient information to monitor the performance of specific subgroups of children, such as LEP children and special education children.

States will decide on the frequency and grade spans for Chapter 1 accountability assessments, except that the assessments, in our view, must include grade 3. (Procedures for assessments below grade 3 are described later.) Assessments should be frequent enough to cover adequately the grades served by Chapter 1 in the state and to provide information on a child's progress at regular intervals, so that children do not fall between the cracks. Annual assessment would be one way to ensure this, although in this case the results should not have to be reported to the federal government more frequently than every other year.

Looking at student outcomes is only one part of a thorough evaluation of the effectiveness of Chapter 1. It is not fair to expect students to perform at a certain level without also ensuring that they receive meaningful opportunities to learn. Therefore, the committee proposes that states develop delivery standards for Chapter 1 programs in prekindergarten through high school.

Delivery standards should address the elements, practices, and inputs that contribute to a high-quality Chapter 1 program and that could be used to determine whether the Chapter 1 program is well designed and effectively delivered in classrooms. As explained later, good delivery standards would be especially important in the early years, because they would be the only standards for assessing prekindergarten and kindergarten Chapter 1 programs under the committee's proposal, and a principal standard for grades 1 and 2. Moving up through the grades, the content and performance standards would take on greater importance in the assessment process.

It is outside the scope and expertise of this committee to specify exactly what the delivery standards should include. Other groups have suggested such factors as a curriculum that meets content standards; instructional practices that can be shown to help disadvantaged students; higher-quality staff development; parental involvement; appropriate and up-to-date materials and equipment; adequate time for teachers to carry out essential tasks; clean, safe, and drug-free schools; availability of educational technology; and availability of social and

nutritional services.[51] Under no circumstances should Chapter 1 dictate how states or local districts allocate any other local, state, or federal resources beyond what Chapter 1 provides. However, states or local districts should have the option, consistent with state law, to set standards affecting their own funding. Delivery standards should allow local flexibility regarding such issues as teaching strategies, instructional materials, and equipment.

As explained later, the state should collect information on delivery standards and use classroom observations to monitor whether these standards are being met.

Reporting and monitoring procedures should be developed by SEAs to assess whether local programs are making sufficient progress toward content, performance, and delivery standards. Under the committee's proposal, SEAs should develop procedures for LEAs to report student performance data and information about program delivery to states for review. SEAs should also develop procedures for monitoring and evaluating how effectively local Chapter 1 programs are being delivered, in accordance with delivery standards. These procedures should include classroom observations of the actual delivery of the program to students. A review by expert practitioners, such as teachers and content supervisors, or by a "state inspectorate" as proposed by other expert panels,[52] could fulfill this role. In states that do not require uniform assessment procedures, state monitoring could include periodic evaluations using state-selected assessment methods.

The data that states receive from the national assessment also should be used, in conjunction with state assessment results, to gauge the effectiveness of Chapter 1 programs and guide state policy decisions.

In developing all these standards and assessments, states could seek assistance from a range of expert sources, which might include universities, Chapter 1 technical assistance centers, and educational research laboratories. States in turn should provide technical assistance to LEAs about state standards, assessments, and procedures.

The committee further recommends different state and local accountability procedures for different age and grade levels, as follows:

Prekindergarten and kindergarten. Although there is interest among early childhood groups, professional associations, and the National Education Goals panel in developing

---

[51]Commission on Chapter 1, *Making Schools Work for Children in Poverty* (Washington, DC: Council of Chief State School Officers, 1992), p. 13. Also, Linda Bol and Edward Haertel, "New Directions Toward the Development of a National Accountability Component in Chapter 1 Evaluation," paper prepared for the Advisory Committee on Testing in Chapter 1, November 1992, p. 7.

[52]Ibid., pp. 81-2.

outcome standards and assessments for younger children based on appropriate levels of child development, this process is in the early stages. At such time as widespread agreement exists on developmentally-appropriate assessments at the prekindergarten and kindergarten levels, these might form a solid basis for Chapter 1 assessment. Until then, the committee recommends that for state and local accountability purposes, prekindergarten and kindergarten Chapter 1 programs be assessed on the basis of delivery standards only, rather than on assessments related to content or performance standards. Thus, neither LEAs nor states would be required to develop content standards, performance standards, or related assessments for Chapter 1 prekindergarten and kindergarten programs; only delivery standards would have to be developed. Furthermore, Chapter 1 would not require norm-referenced testing or other testing of children at these levels. This would not preclude state or local districts from conducting their own assessments of young children if they desire.

It is outside the charter of this committee to specify the delivery standards for prekindergarten and kindergarten, but they should be high and reflect the best research about early childhood education. We recommend that standards be developed by state panels with expertise in early childhood programs. Evaluations of how well delivery standards are being met should include observations of classroom practices.

Grades 1 and 2. In recommending procedures for Chapter 1 assessment at grades 1 and 2, the committee faced a dilemma. On one hand, if school districts do not conduct any assessment for accountability purposes of children at these ages--who comprise a significant percentage of Chapter 1 children--they run the risk that young children may not be served or may be served ineffectively. On the other hand, if districts do standardized testing at these grades for accountability or student selection purposes, they may run the risk of driving instruction in counterproductive ways.

The ultimate answer will emerge in the long run, as the work of National Education Goals panel #1 on school readiness proceeds and as a consensus on standards for child development and content knowledge is developed for these grades. Until then, the committee recommends that LEAs conduct some type of content assessment for grades 1 and/or 2. The content assessments should be congruent with state-developed content and performance standards for grades 1 and 2 (or locally developed ones, where the state permits), and should lead into the state content and performance standards and related assessment for grade 3. Assessments for Chapter 1 testing at grades 1 and 2 should be developmentally appropriate and performance based, and should include multiple methods such as trained observations, portfolios of student work, demonstrations, and developmental checklists.

We further recommend that states evaluate first- and second-grade programs in terms of how well they are implementing delivery standards, again using classroom observations.

The committee does not intend that all children in grades 1 and 2 be assessed for purposes of state or local accountability, nor do we intend that the state aggregate data from these grades. The assessments should include some measures of oral language for LEP and

special education children at these grades.

Grades 3 through 12. For these grades, students and programs would be assessed on the basis of content standards, performance standards, related assessments, and delivery standards. State assessment practices for grades 3 through 12 should accurately appraise the types of knowledge and skills and the levels of proficiency expected at various grades or stages. Assessments must include multiple measures that are valid, relevant, and include alternative assessment, such as demonstrations of performance. Tests results could be normed, if that was desired, although the main purpose would be to assess the performance of students against standards, not each other.

Special populations. States and local districts should take steps to ensure that LEP students and special education students who are eligible for Chapter 1 are included in assessments for accountability purposes. It is important to know whether these children are progressing in Chapter 1.

LEP students should be assessed in a way that appropriately measures their achievement. Specifically, LEP students should be assessed in terms of oral language skills and written language skills, for both accountability and eligibility purposes.

Special education students present a different set of challenges for Chapter 1 assessment. Appropriate assessments should be administered in the right environment to hearing-impaired, visually impaired, and orthopedically impaired students. With respect to learning disabled (LD) children in Chapter 1, several complex issues arise that go well beyond assessment: Under what circumstances should LD children be eligible for Chapter 1? What should be done about children who have not yet been diagnosed as LD but are obviously having difficulties learning? To what extent are educational difficulties attributable to a learning disability, as opposed to other indicators of educational deprivation? Because of the complexity of these issues and the limited time available for deliberations, the committee was not able to debate these issues thoroughly. More work needs to be done by a panel of experts in the special education field.

Secretary's approval. States must submit to the Secretary of Education for review and approval plans that include the content standards, performance standards, delivery standards, assessments, and monitoring and reporting procedures. The Secretary would establish sound criteria for reviewing state plans. As a starting point, the committee suggests several broad criteria.

Content and performance standards for Chapter 1 children should be:

- substantive, demanding, and ambitious, reflecting high expectations for students;
- the same as the standards for all students;

- absolute rather than relative; in other words, they should specify the knowledge and skills students are expected to attain, rather than compare students with each other;

- progressive; in other words, the standards at one level or grade should be related developmentally to those at the next level and should specify what students should know and be able to do at different points in their educational careers. States should have the flexibility, however, to develop different types of progressive benchmarks rather than just one;

- equitable between Chapter 1 and the regular program.

State assessments should:

- validly measure the content and performance standards;

- be discernibly linked to the state standards. Using the same tests now used for Chapter 1 evaluation in the same way would not suffice. Nor would it suffice for states to select their standards merely because they conform with existing state tests; rather, the process should begin with the standards, then move to the assessments. Similarly, the Secretary should first appraise the standards, then examine whether the assessments are appropriate to those standards;

- meet high professional and technical measurement standards;

- include criteria for assessing local progress toward standards, which may include a requirement for LEAs to submit some type of data annually.

As regards delivery standards, the Secretary should review them only to ensure that they are within the general guidelines for Chapter 1 but should not have authority to regulate additional standards. The Secretary should also approve the state's proposed reporting and monitoring system and ensure that the state plan provides for adequate reporting of information to the Secretary, including the information the Secretary would need to conduct the review of assessment data from prekindergarten through second- grade programs (see Recommendation #1).

The Secretary should not develop a single national model for content or performance standards or assessments. The Secretary could, however, compile several examples of sets of state standards, assessments, and procedures that meet approval criteria.

The committee does not intend that state plans be set in concrete. Rather, we see them as dynamic and subject to revision as states gain more knowledge and experience about standards and assessments, and as consensus builds about voluntary national standards.

The Secretary should appoint and rely on a broadly representative, independent panel

of experts to assist in developing criteria and reviewing state plans against these criteria.

Local Accountability

Recommendation #3. To ensure accountability for federal funds, local education agencies should assess the progress of Chapter 1 children in grades 3 through 12 using state-developed assessment methods (or locally adapted ones where states permit) tied to content and performance standards for all children. Different procedures should be used to assess programs for prekindergarten and kindergarten and for grades 1 and 2.

The committee's proposals for local accountability are closely linked to those for state accountability. As explained in Recommendation #2, state standards and assessments would guide accountability at the local level, unless the state chose to permit LEAs to modify state standards and assessments or to develop their own subject to state review and approval.

Specifically, the committee recommends that each LEA submit a plan to the SEA that describes how it will implement or adapt state standards and assessments in its schools. The LEA plan should also describe how Chapter 1 services will be provided in accordance with delivery standards, and may also note school-by-school variations or exceptions based on different populations or needs.

The results of local accountability assessments should include multiple measures and should be capable of being aggregated by school or district, so that it would be possible to determine what proportion of Chapter 1 children are meeting performance standards. Local assessments should also provide adequate information to determine whether Chapter 1 students are maintaining progress as they move through the higher grades. Sampling techniques could be used in districts that have enough children to ensure a valid sample.

LEAs should report assessment results and descriptive information about delivery standards to the state, according to a state-determined schedule and format (which would not have to be annual). As discussed in the previous section, states should have procedures for monitoring student performance and evaluating service delivery at the local level.

It is not within the mandate of this committee to define a new system of program improvement. However, if Congress amends Chapter 1 to continue something akin to the program improvement requirements, the state and local accountability procedures we are recommending could form a basis for determinations about the effectiveness of individual school programs, including schoolwide projects, and the progress of individual students. The important point in the committee's view is that such determinations should be based on multiple measures, and the system we recommend would provide this option. It is also recommended that LEAs conduct their own formative evaluations of the quality of Chapter 1 programs in individual schools, whether or not the law maintains program improvement.

Finally, the committee recommends that legislation and regulations be revised to eliminate the provision that ties the amount of funding a school receives to its test scores, thereby creating a negative incentive for improved student performance. A funding formula more tightly tied to poverty would seem to be a fairer approach.

Instructional Feedback and Diagnosis

Recommendation #4. Teachers should assess Chapter 1 student learning and diagnose student needs using continuous, intensive, and varied methods and should use the feedback from these assessments to plan and adjust instruction.

Chapter 1 teachers and other school professionals currently use a range of formal and informal assessment methods to monitor student learning, diagnose student educational needs, and inform their own teaching. This type of ongoing classroom assessment is critically important to Chapter 1, but receives only limited emphasis in the existing framework of Chapter 1 testing requirements. The committee seeks to give greater recognition to this function of assessment and to eliminate any disincentives in current law that discourage teachers from using a range of assessment methods for this purpose.

The committee recommends that Chapter 1 law and regulations explicitly endorse the use of continuous, intensive, and varied assessments for instructional feedback and diagnosis at all age and grade levels. Multiple assessments, including performance assessment, should be used.

It is important to clarify the relationship between assessment for instructional feedback and assessment for local and state accountability purposes. The committee envisages assessment for instructional feedback as an ongoing process, controlled by the teacher and embedded in what teachers regularly do. The purposes would be to monitor and reinforce student learning and help teachers plan their next instructional steps. Thus, the results of periodic state and local assessments for accountability purposes would be just one source of instructional feedback. Other sources might include teacher-initiated assessments, informed teacher judgment, classroom observations, performance assessments, computer-adaptive testing, and more.

Teachers should be encouraged and supported to select and implement classroom assessments that would best enable them to follow students' progress toward content and performance standards. For example, in the interest of encouraging higher-order thinking skills, Chapter 1 students might work on a challenging demonstration or be encouraged to think about how they arrived at an answer to a problem. States and LEAs could support teachers in these practices by providing exemplary items or tasks or creating banks of activities for classroom assessment that are related to their instructional goals.

With regard to assessment for diagnostic purposes, the committee recommends that diagnosis of Chapter 1 students' educational needs be undertaken by teachers as they observe

children in action in the classroom and that diagnosis focus on identifying students' strengths and knowledge. Districts should feel free to choose from the full range of diagnostic assessments--diagnostic tests, developmentally appropriate assessments, portfolios, checklists, documented teacher observations, and oral language assessments, for instance.

Not all school professionals have been fully prepared to conduct the range of assessments envisaged for instructional feedback. Thus, staff development, as discussed below in Recommendation #8, is absolutely essential.

In recommending continuous and varied classroom assessments, the committee does not intend for districts to burden teachers with too much recordkeeping or to require teachers to document or aggregate the results of every classroom assessment. Neither is it our intent to compel teachers to implement more classroom assessments than they reasonably need.

Even so, in the course of a school year, the classroom assessments used by teachers would produce a great deal of information about the progress of Chapter 1 children across the school. The committee encourages schools to review this information as part of their self-examination processes, in a way that is not burdensome or unfair to teachers.

Student Eligibility and Identification

Recommendation #5. School districts should use multiple indicators to identify students with the greatest needs for Chapter 1 services, including special procedures for young children. Appropriate methods should be used to identify and select limited-English-proficient students and special education students who are educationally disadvantaged for inclusion in Chapter 1 programs.

Wherever possible, the committee recommends that high-poverty schools adopt a schoolwide project approach (see Recommendation #6). Because all students in a schoolwide project may benefit from Chapter 1-funded services and activities, this approach eliminates incentives to openly identify and separately group Chapter 1 students.

In schools that cannot or do not adopt a schoolwide model, the committee recommends continuing the concept of selecting children with the greatest educational needs for Chapter 1 services. To identify these children, school districts should use multiple indicators at all grade levels.

The committee would require all districts to use multiple forms of appraisal to augment or replace the use of a single indicator. In addition, as explained later, the committee recommends that districts use different types of criteria for prekindergarten through grade 2, and for grades 3 through 12.

Informed professional teacher judgment--intended to mean something more precise, objective, and systematic than intuition--should be a component of student selection at all

levels, along with performance-based or other empirical assessment results. Whereas the committee is aware of the possibility that teacher judgment is subject to bias, the identification process should give credit to the judgments of those professionals who serve most closely with children and are most familiar with children's educational needs. As the experience of selected states and districts demonstrates, teacher judgment can be systematized to lessen concerns about objectivity.[53]

Sensitivity to the different developmental stages of children is called for in selection. For children at the prekindergarten level through grade 2, the committee recommends that districts select children for participation in non-schoolwide Chapter 1 programs primarily on the basis of family poverty. In addition, school districts should consider other factors that may place young children at educational risk. Illustrations of factors that might be considered include low birth weight or other biological risk; a diagnosed medical disorder; parents who are themselves educationally disadvantaged; being born to a teenage mother; evidence of emotional deprivation; residence in a home where English is not the primary language; residence in a home where there is substance abuse, child abuse, or chronic mental illness; or other special circumstances, including homelessness or placement in foster care. These are not the only risk factors, of course, nor do all of these factors create educational risk for every child. In providing these examples, the committee does not intend for school districts to collect new information about children's health, parental situations, or home circumstances. Rather, we encourage schools to consider the information about children in prekindergarten through grade 2 that is already available in school records or from teachers. In any case, LEAs should apply informed teacher judgment to decisions about eligibility and identification for children at the prekindergarten level through grade 2, including judgments about children's oral language development.

For children in grades 3 through 12, the committee recommends that programs that are not schoolwide identify children using multiple, objective, and uniformly applied criteria, increasingly oriented to state and local content and performance standards as these standards become available. These criteria could include such measures as state-developed tests aligned with state standards, norm-referenced tests, criterion-referenced tests, alternative assessments, student classroom performance, teacher judgment, proficiency in oral language, and others. These criteria should include assessments of proficiency in reading, writing, oral language, and mathematics, and may include other subjects addressed by state standards and assessments.

At all age and grade levels, the committee recommends that school districts develop special means to ensure that eligibility and identification methods are applied appropriately to LEP students and special education students, and that eligible children from these groups are included in Chapter 1 programs.

---

[53]Edward Reidy, Kentucky Department of Education, presentation to the Advisory Committee on Testing in Chapter 1, September 28, 1992.

For LEP students, several points warrant attention:

- Oral language skills, as well as written language skills, must be assessed as part of the selection process.[54] Because of the connection between oral and literacy skills among LEP students, it is necessary to assess reading, writing, listening, and speaking skills.[55]

- Many LEP students, though limited in English language skills, are highly fluent in the home language, to the extent that schools or states could offer instruction in the home language as a complement to English language instruction.

- It should be borne in mind that oral language problems exist for majority as well as language minority students, and these problems are of concern for all students.

As already noted, eligibility and identification policies should not be used to exclude special education children. Because the committee was unable to consider all the complexities of this issue, however, we recommend further study by an expert panel.

A final issue related to eligibility is the question of when a child is ready to leave Chapter 1. Under the existing system, children could have breaks in Chapter 1 service as they "test out," in other words, if they score above a certain cutoff on a standardized test. Under the committee's paradigm, a child would be eligible until he or she achieves proficiency according to state or local performance expectations and related assessments.

Schoolwide Projects

Recommendation #6. Chapter 1 assessment should be compatible with schoolwide approaches to service delivery, as well as with other effective

---

[54]The Council of Chief State School Officers has analyzed state assessment and data collection practices for LEP students and developed recommendations for improving assessment for these children. The Committee suggests that Chapter 1 administrators examine these practices when developing and implementing assessments of oral and written language for LEP children. See Oona M. Cheung and Lisa W. Solomon, *Summary of State Practices Concerning the Assessment of and the Data Collection about Limited English Proficient (LEP) Students* (Washington, DC: Council of Chief State School Officers, 1991); and Council of Chief State School Officers, *Recommendations for Improving the Assessment and Monitoring of Students with Limited English Proficiency* (Alexandria, VA: Weber Design, 1992).

[55]Many students from monolingual English backgrounds may also have insufficiently developed oral English language skills to participate fully in "mainstream" classrooms; these children may benefit from similar consideration.

models for Chapter 1 instruction.

Although it is not the committee's charge to make recommendations about service delivery, our analysis of assessment practices leads us to conclude that schoolwide projects are a highly desirable option for delivering Chapter 1 services. Schoolwide projects address some critical assessment dilemmas in Chapter 1; for example, in a schoolwide model, there is no identification or separate grouping of Chapter 1 students for instructional purposes. Moreover, they are compatible with the committee's hope to transform Chapter 1 schools into learning organizations. When properly implemented, the schoolwide project model also fits well with our paradigm of high expectations for all children. Furthermore, schoolwide projects can be an efficient way for the poorest schools to use resources. However, the benefits of a schoolwide approach accrue only when schools plan meaningful changes in school processes and instructional practices.

Accountability for the progress of disadvantaged children continues to be important in schoolwide projects. Thus, the committee recommends that states and local districts develop means for determining whether the most educationally disadvantaged children within schoolwide projects--the children who in another context would be selected for Chapter 1--are making sufficient progress. It is recommended that schoolwide projects continue to identify and follow the progress of Chapter 1-eligible children and other subgroups of children, such as LEP children, using multiple measures that conform with state and local assessment plans. If students who would traditionally be eligible for Chapter 1 are not making adequate progress, the school would need to take steps to ensure that they do.

The committee recommends that states and local school districts use the transition period to develop ways to identify the most educationally disadvantaged children in a schoolwide context and to define what constitutes an excellent schoolwide project.

Of course, schoolwide projects are not the only effective model for Chapter 1 services. Extended-day programs, in-class models, and other approaches can be equally effective. The assessment system we propose would be appropriate for all effective models.

Transition Period

Recommendation #7. The five years following the forthcoming reauthorization of Chapter 1 should serve as a transition period, with new elements phased in as they become ready. By the end of the five years, state standards, assessments, and procedures for Chapter 1 accountability should be in place, and a national assessment for Chapter 1 should be operational. During this transition, Chapter 1 assessment should operate according to state transition plans, approved by the Secretary, that include multiple assessment methods.

The committee realizes that its vision of Chapter 1 assessment includes several components that have not yet been developed or fully tested. In most content areas,

consensus does not exist about voluntary national standards; the efforts under way to develop them are not due to be completed until 1995. Many forms of alternative assessment are still in the experimental stages, with refinements needed before they can be implemented widely in a program as important as Chapter 1. States differ greatly in both their expertise about assessment and their applications of content and performance standards. In short, many policy steps, and not a few technological ones, remain to be taken before a fully integrated system of instruction, standards, and assessments will become a reality in the nation.

A change of the order proposed in this report will take time. The current Chapter 1 evaluation system is too interdependent and well established to be transformed overnight. Many people at the local, state, and federal levels and in the private sector must be involved, with support from universities, laboratories, research centers, professional associations, and many others. Fundamental research must be conducted, new organizational agreements must be forged, and new technological systems must be developed.

Therefore, we propose a five-year transition period, to begin upon enactment of new amendments to Chapter 1. We further suggest that this period be divided into a two-year phase and a three-year phase, with several steps proposed for each phase.

Immediately upon enactment, the Secretary should propose regulations to govern the transition, and soon afterward states should submit transition plans describing how they will ensure accountability during the interim period, consistent with the recommendations explained below.

In addition, during the first two years, states would develop content standards, performance standards, and delivery standards for Chapter 1 and submit them to the Secretary of Education. By the end of the first two years, the Secretary should have completed approval of all state standards. This means that the Secretary would have to move quickly after enactment of new amendments to develop regulations and approval criteria and appoint an advisory panel.

During the last three years, states would develop and submit Chapter 1 assessment plans and instruments, with the Secretary again completing the review and approval process before the end of the transition period. Thus, by the end of five years, all states would have approved standards and assessments ready to be implemented or already being implemented.

We envisage this process as staggered. States that reach consensus quickly should submit their plans early and begin implementing them as soon as they are approved. Other states may need the full five years. We also acknowledge that there will be timing complexities and obstacles, and that the transition process will have to be fluid. For example, during this time, professional associations and related groups will continue their efforts to develop national standards. Where national standards already exist, as in the case of mathematics, states should use them to guide development of content and performance standards. Where voluntary national standards do not exist, the urgency of Chapter 1 testing

reform makes it important for states to move ahead regardless.

During the transition, we recommend that the federal government support and lead an intensive research and development effort and a staff development effort, as further explained in Recommendation #8 below. The federal government should also provide technical assistance to states and local districts to help solve the logistical challenges of a new Chapter 1 assessment system. Administrators, teachers, and parents would be important participants in technical assistance. The committee further suggests that technical assistance should address ways to implement the new provisions with the least amount of paperwork and administrative burden and at a reasonable cost.

The Secretary should also move quickly to implement national accountability within five years. Which procedures will govern Chapter 1 assessment during the transition? The committee advises against suspending evaluation requirements altogether and creating an information vacuum; we also advise against maintaining the current system unchanged for another five years. A reasonable option is for the Congress or the Secretary to develop guidelines for the transition along the following lines:

- States and LEAs would continue to administer some form of assessment for accountability purposes, including student accountability or a version of program improvement; however, they should use multiple assessment methods, not just the norm-referenced tests now used. If a state testing program is available to fulfill this need, it should reach an adequate number of Chapter 1 children and grades.

- The requirement to aggregate TIERS data at the national level should be discontinued for the transition.

- LEAs should not be required to aggregate assessments unless the state requires it.

- The Secretary could require states to report some form of assessment results; whether they would be aggregated would be up to the state.

- So that there will not be an information vacuum at the federal level, the federal government should conduct a national evaluation on a sampling basis, perhaps using methods and procedures like those in the Prospects longitudinal study.

- New components of the assessment system should be phased in as they become ready. For example, states might pilot-test some assessments in selected districts or subjects, then implement them on a wider scale. States are also encouraged to share information and to learn from each other.

- If Congress continues some version of program improvement, states and local districts should use multiple measures to make these types of decisions during the transition and beyond.

Federal Government Development Assistance

Recommendation #8. To help teachers, state and local administrators, and other key professionals implement these recommendations, the Chapter 1 legislation should include a funding set-aside for staff development related to assessment and performance standards. The federal government should also support a national effort to expand and refine the knowledge base about assessment and standards.

Staff development is vital to the success of all the recommendations contained in this report. Therefore, the committee recommends that in reauthorization, Congress provide a set-aside of funds specifically for staff development to help states and school districts implement standards and related assessments for Chapter 1.

Staff development should provide teachers, state and local administrators, and other Chapter 1 personnel with the skills, knowledge, and attitudes that will enable them to:

- develop and implement high-quality assessments tied to standards and built upon the most effective practices from research;
- use alternative assessments fairly and appropriately;
- align assessments with new standards and with other school testing programs;
- make better use of feedback from multiple assessments to plan and adjust instruction.

In carrying out staff development, the federal government should consider the resources available from such sources as colleges and universities, state and local education agencies with exemplary assessment programs, and the private sector.

The Secretary should ensure that the staff development supported is of the highest quality possible and incorporates the best practices from research in a way likely to make real improvements in the classroom environment.

The federal government should also lead and fund a significant research and development effort. Several elements must be included. More research and demonstration projects must be conducted to ensure that alternative assessments meet professional standards of reliability, validity, and generalizability and can be implemented with reasonable cost and efficiency. Further research must also be done to ensure that alternative assessments are culturally sensitive and fair. Attention must also be paid to how alternative assessments could be connected to curricular frameworks and standardized for useful, wide-scale implementation by teachers, and whether those assessments are appropriate for such decisions as Chapter 1 selection or program improvement. Other issues for R&D include appropriate assessments for young children and the applications of technology to assessment and

management. The role of technology for managing, as well as administering, assessment should also be explored.

Research and development funded under this effort should build upon the body of research knowledge that already exists about effective assessment. Research could be carried out through partnerships of SEAs, LEAs, universities, and the private sector.

States and local districts should be intimately involved in this effort at all stages, because many are already on the cutting edge of assessment and they will be the agencies to implement the results. The development process should also include input from people at the school level who will use the assessments--teachers, administrators, and parents.

The private sector will be especially important partners in this R&D venture. Testing companies should be encouraged to cooperate with public agencies.

CONCLUSION

Changing the paradigm for Chapter 1 testing will require leadership and cooperation from the federal government, SEAs, LEAs, practitioners, researchers, education professional organizations, and others. It will also require federal funding for such critical activities as research, development of standards and accompanying assessments, technical assistance, and professional development.

Some might argue that it is unrealistic to expect such a transformation to take place in five years. We respond by noting that the whole history of Title I and Chapter 1 has been characterized by people accomplishing remarkable things in a relatively brief time. Title I proceeded from draft bill to law in a matter of months. A year later $1 billion was made available to school districts. Within a few years, Title I had completely transformed the way people thought about the needs of disadvantaged children. If that can be accomplished, so can a reform of Chapter 1 testing aimed at improving program quality and student learning.

VIEWS FROM INDIVIDUAL COMMITTEE MEMBERS

Edward A. De Avila
Linguametics Group
Oakland, California

While I am in agreement with most of the principles as outlined in this document, I still have serious reservations. I have presented these reservations in greater detail in other written comments. The following summarizes what I have said before:

1. I agree in principle with the idea of requiring the states to do more in defining "standards" and assessing educational growth. I do not feel, however, that the draft provides sufficient detail or assurances that states will be held accountable to do the "right thing," rather that they will be left to "do their own thing." This places us squarely back were we were at the inception of Chapter 1 (Title I) in 1964. Instead of one program we could just as well end up with fifty different programs.

2. I agree (with almost everybody in the western world) that NRTs, in general, are perhaps not the best way to assess student progress and that there is a need to shift toward a more performance based assessment of what is learned. This is particularly true at the lower grades. However, until such time as alternatives are fully developed and in place it strikes me as foolhardy to argue or imply that all forms of "standardized" testing be discontinued.

   We should be mindful of the current problems and limitations with "authentic" portfolio assessment. For example, the recent RAND report on Vermont indicates that "reliability" on these measures is not so readily obtained. Similarly, California (CAP), which is perhaps one of the most advanced states in developing "authentic" assessment, has made few, if any, provisions for students who are of limited English proficiency.

3. I agree that there are developmental differences that impact test performance and that the continued use of NRT's at the lower grades is folly. In this connection, however it is important to note that a shift toward "absolute performance standards" does not in itself preclude the use of standardized tests in general or NRTs in particular. It only refocuses the metric from one of relative status (i.e., percentile rank and NCE) to one of absolute proficiency (i.e. proficiency levels).

   Moreover, I do not agree with the idea that all accountability below the third grade should be restricted to measures of the extent to which programs have met "delivery standards," particularly if no such standards exist! There are a good many measures currently available that are useful below the third grade. A thoughtful review of these instruments would certainly be helpful.

4. I agree that "steps should be taken to ensure appropriate assessment" of language minority (and handicapped) students. Without more detailed consideration for the changing demographics and its impact or the linguistic make up of the Chapter 1 eligible population, the statement is virtually meaningless.

   This seems a cruel disregard for the millions of language minority students that will soon make up most of the eligible Chapter 1 population. Moreover it denies the increasing importance of language proficiency (Reading, Writing, Listening and Speaking) for all students, not just "LEP's," particularly, below the third grade!

5. I agree that "national accountability" can best be accomplished by means of a national "NAEP-like" assessment conducted on a periodic basis. However, I am not comfortable with the failure of the document to more clearly define what is meant by this recommendation and how it would play itself out in reality.

   I feel that the document would have been more constructive had it put more focus on these issues rather than on restating the all too obvious and hackneyed criticisms of norm referenced tests.

   In summary I feel that the recommendations as stated in this document will shift what has been the federal responsibility to states that are poorly equipped or motivated to take on the additional burden. The new effect will be to replace Chapter 1 as a national program with fifty different locally defined programs. One can only guess as to where this will lead.

Robert B. Frary
Office of Measurement and Research Services
Virginia Polytechnic Institute and State University

The Report of the Advisory Committee on Testing on Chapter 1 is flawed in its conclusion that norm-referenced, objective tests should not be used at all for evaluation of Chapter 1 in the first and second grades. The report goes to great lengths to point out the inadequacies of these tests for diagnosis and placement of students and for provision of short-term instructional feedback. The fact that they may be developmentally inappropriate for some (but certainly not all) younger examinees is argued at length. It also points out the inadequacy of the present (TIERS) scheme for using norm-referenced tests for evaluation of Chapter 1 outcomes, which is actually a misuse of norm-referenced scores. However, these facts cannot logically be used to conclude that objective tests should not be used, even as early as the latter part of first grade, for program evaluation, or that the results could not be reported and used in an effective and constructive manner. It is important to understand in this regard that, if a student is developmentally unready or otherwise unable to deal with some types of test questions, the test can nevertheless correctly report the student's lack of achievement. At the same time it can correctly evaluate the success of educational programs

in teaching skills and knowledge to students who are sufficiently mature to deal with the material. Of course, a test that an examinee simply cannot understand should never be administered, and, in the lower grades, some students are not able to deal properly with objective tests. But objective tests can validly document the achievement levels of large proportions of students in the primary grades as evidenced by their successful use for generations. The report buttresses its negative view of norm-referenced, objective tests (not only for grades one and two, but for all Chapter 1 evaluation) by suggesting their replacement with alternative approaches such as performance tests, portfolio analysis, etc. These methods have the potential for providing useful information about students that may not be available from objective tests, but they are cumbersome and expensive to apply, and their psychometric characteristics (reliability and validity) have typically been found to be inadequate for evaluation studies. To expect these types of measures to be adequate replacements for norm-referenced, objective tests in the near or even long-term future for purposes of program evaluation at any level is, to say the least, an example of extreme optimism.

Brenda Joyce Lee
Principal, Glenallan Elementary School
Montgomery County (Maryland) Public Schools

Considering the national trend, it may be advisable to change the word handicapped to disabled. My interpretation of the Individual Disabled Educational Act (IDEA) indicates that the word disabled is the correct term.

I would also like to suggest that we elaborate and place emphasis on the limited English speaking student (ESOL) in relation to effective ways of assessing these students.

Milton D. Matthews
Director, Division of Compensatory Education
Mississippi Department of Education

The report addresses a state-by-state assessment to establish a baseline for national accountability. I concur with this procedure. However, I believe that, when state data are available as part of the national sample, states should compare their state evaluation results with national state sample results. While I recognize the cost factor that is inherent in providing state data on each national assessment, I believe that there is some off-set to be gained when one considers that mobility is a factor in today's global society. States should be required to report comparisons of state data to national data to the Secretary on a regular basis.

Diane J. Sawyer
Office of Dyslexic Studies
Middle Tennessee State University

Early identification and intervention has long been a frequently voiced objective for the academically at-risk child. The report of this Advisory Committee on Testing in Chapter 1 is our opportunity to indicate how Chapter 1 assessment can provide a vehicle to implement that objective and to document the long-term impact of early intervention. We should not miss this opportunity.

It is not clear to me how voluntary standards will relate to the content of the NAEP-like periodic evaluation of progress in Chapter 1. Will this tool not effectively mandate adherence to specific standards in the various content areas? Further, since earlier in the report a specific comment was made regarding the potential for revisions in Chapter 1 to effectively transform education in the United States, it seems that this later statement, "It is not intent of the committee to undermine the voluntary nature of any national standards, however," is probably not accurate.

Robert S. Stephenson
Office of Educational Accountability
Dade County (Florida) Public Schools (retired)

My comments essentially focus on an unacceptably low level of accountability for the Chapter 1 program at grades two and below, including pre-K. In pre-K and K, sampling-based individual assessments should be acceptable if conducted by appropriately trained persons. At grades one and two, this individual assessment strategy might be continued, but I know of no valid reason (or research) indicating that NRTs are harmful to students at these grades. The "harmful" reference is, of course, to an assumption apparently underlying the lack of accountability below grade three. And, given that grades two and below include the intervals where cognitive development is most rapid, I cannot agree with a position that precludes the use of standardized instruments to gather program impact data at the (sampled) student level in this most important age interval.

Karen Kring Wixson
School of Education
University of Michigan

National Accountability in Grades 1 & 2. I would favor changing the statement regarding standardized performance data to include grade 1 as well as grade 2. I felt that the arguments made to the Committee by individuals like Robert Slavin were compelling in terms of the need for national accountability at lower grade levels. I would, however, provide for teacher judgment in determining whether individual students be given all or part

of the assessment, and require that statistics be provided on the number of students deemed unprepared to take all or part of the assessment.

APPENDICES

APPENDIX A

CHARTER

ADVISORY COMMITTEE ON TESTING IN CHAPTER 1

UNITED STATES DEPARTMENT OF EDUCATION

THE SECRETARY

CHARTER

ADVISORY COMMITTEE ON TESTING IN CHAPTER 1

AUTHORITY

20 U.S.C. 1233a. The Advisory Committee on Testing in Chapter 1 (Committee) is governed by the provisions of Part D of the General Education Provisions Act (P.L. 90-247, as amended; 20 U.S.C. 1233 et seq.) and the Federal Advisory Committee Act (FACA) (P.L. 92-463; 5 U.S.C.A. Appendix 2), which set forth standards for the formation and use of advisory committees.

PURPOSE AND FUNCTIONS

The "1992 National Assessment of Chapter 1 Act" (P.L. 101-305, Sec. 2(b)(3)(C)(ii); 20 U.S.C. 2882 note) requires the Secretary of Education (Secretary) to assess the "adequacy of standardized tests" used to measure the academic achievement of Chapter 1 students. To enable the Department to provide the Congress with information in this area, the Advisory Committee on Testing in Chapter 1 will conduct a comprehensive examination of the standardized tests currently used and will advise the Secretary on possible improvements or alternatives for current testing procedures in the Chapter 1 program. These recommendations shall include possible changes in test instrumentation, administration, and the reporting of test results at the State and local levels.

The Secretary is required to develop national standards for local evaluation of programs under Public Law 100-297, Section 1435 (20 U.S.C. 2835). The statute permits the Secretary to use the Title I Evaluation and Reporting System as a model. This model relies exclusively on norm-referenced testing and reporting on a scale of normal curve equivalents. Such a model is necessary for the Secretary to meet his reporting requirements to the Congress (Sec. 1435(b), which necessitate aggregating data across school buildings, districts, and States. The Committee will provide guidance on developing national standards that may permit the aggregation of data and the measurement of program impact without depending exclusively on norm-referenced testing.

400 MARYLAND AVE. S.W. WASHINGTON D.C. 20202-0100

The Committee will also provide recommendations to the Secretary on regulatory or legislative changes that will allow an alternate assessment procedure which must provide information upon which schools, districts, and States can be held accountable for student performance. Such alternatives should also permit national aggregation of student performance measures. Alternative assessment should feature as many of the following characteristics as possible:

It should take advantage of State and locally developed tests and advances in assessment based on student performance and testing that measure mastery of learning objectives or other educational criteria in the basic and advanced skills.

It should present information in a format that is intuitively understandable to the general public.

It should be able to assess performance of limited English proficient students and children with handicapping conditions.

It should provide information that can be used to improve and reform instruction, curriculum, and school operations.

It should be able to be implemented with reasonable cost and minimal administrative burden.

STRUCTURE

The Committee consists of not more than 17 public members appointed by the Secretary. The Secretary designates a chairman and a vice chairman from among the members.

The Secretary appoints the membership of the Committee after consultation with appropriate educational organizations and other interest groups. Members include individuals who have extensive backgrounds in educational evaluation, assessment and testing, education for the disadvantaged, educational administration, including local school boards, and parent involvement in schools, and represent a broad range of viewpoints and experience. The term of office of each member is 2 years, or for the life of the Committee, whichever is shorter. Any member appointed to fill a vacancy serves for the remainder of the term for which the predecessor of such member was appointed.

The Committee may establish subcommittees composed exclusively of members of the committee. Each subcommittee complies with the requirements of applicable statutes and regulations. Each subcommittee presents to the Committee its recommendations for subsequent action by the full Committee. Timely notification of each subcommittee establishment and changes, including its charge, membership, and frequency of meetings is made in writing to the Committee Management Office. All subcommittees act under the policies established by the Committee as a whole.

Management and staff services are provided under the direction of the Designated Federal Official (DFO) who is appointed by the Assistant Secretary for Elementary and Secondary Education.

MEETINGS

The Committee meets approximately six times during its existence at the call of the Secretary or Designated Federal Official (DFO) who approves the agenda and is present at all meetings. Standing committees meet as required at the call of their chairperson with the concurrence of the Committee Chairperson and the DFO. Meetings are open to the public except as may be determined otherwise by the Secretary of Education or designee in accordance with Section 10(d) of the Federal Advisory Committee Act (FACA).

Adequate notice is given to the public in advance of all meetings.

Meetings are conducted and records of the proceedings kept, as required by applicable laws and departmental regulations.

ESTIMATED ANNUAL COST

Members who are not full-time Federal employees are paid at the rate of $100 per day, plus per diem and travel expenses in accordance with Federal Travel Regulations.

Estimated cost of operating the Committee, including compensation and travel expenses for members, but excluding staff support, is $150,000. Estimated person-years of staff support is one-quarter at a cost of $15,000.

REPORTS

The Committee will issue a final report, including its recommendations, to the Secretary by one year from the date of its initial meeting. In addition, the Committee makes a report of its activities, findings and recommendations to the Congress not later than March 31 each year. The report contains as a minimum a list of members, their business addresses, the Committee's functions,the dates and places of meetings, and a summary of the Committee's actions and recommendations made during the prior fiscal year. These reports are included with the Secretary's Annual Report to the Congress and are also submitted to the Committee Management Officer.

TERMINATION

The Committee terminates one year from the date of its creation unless the Secretary determines in writing not more than thirty days prior to the expiration of such year that the Committee's existence for an additional period, not to exceed one year, is necessary in order to complete the recommendations or report for which it was created. This Charter expires two years from the date of filing, in accordance with Section 14(a)(2) of FACA. The Committee is hereby chartered in accordance with Section 9 of FACA.

APPROVED:

1/29/91
Date

Acting Secretary

Filing Date: December 1, 1991

UNITED STATES DEPARTMENT OF EDUCATION

THE SECRETARY

ADVISORY COMMITTEE ON TESTING IN CHAPTER 1

I hereby determine, after consultation with the Director, Committee Management Secretariat, that the establishment of the Advisory Committee on Testing in Chapter 1 is necessary and in the public interest in connection with the performance of duties imposed on the Department by law, and that such duties can best be performed through the advice and counsel of such a group.

1/24/91
Date

Acting Secretary

400 MARYLAND AVE. S.W. WASHINGTON, D.C. 20202-0100

APPENDIX B

PAPERS COMMISSIONED BY THE ADVISORY COMMITTEE

Bol, Linda, and Edward H. Haertel. "New Directions Toward the Development of a National Accountability Component in Chapter 1 Evaluation." November 1992.

Estes, Gary D. "Reflections on Testing in Title I/Chapter 1." April 1992.

Hawley, Willis D. "Using Chapter 1 To Fundamentally Revision Schools and Learning." September 1992.

Pfannenstiel, Judy. "Measures of Achievement in Chapter 1." September 1992.

Schenck, E. Allen. "Summary of Recommendations Regarding Chapter 1 Assessment from Selected Sources." September 1992.

Timar, Thomas. "Federal Educational Policy and Practice: Building Organizational Capacity Through Chapter 1." September 1992.

ACKNOWLEDGMENTS

We are grateful to the following individuals for their presentations to the Advisory Committee:

Francie Alexander
Office of Educational Research and Improvement

Richard Allington
The State University of New York, Albany

Linda Bol
Art, Research & Curriculum Associates

Jacqueline Cheong
University of California, Berkeley

Katherine Conner
The School District of Philadelphia

Gary Estes
Far West Laboratory

Michael Feuer
Office of Technology Assessment

Steven Gorman
National Center for Education Statistics

Edward Haertel
Stanford University

Willis Hawley
Vanderbilt University

Asa Hilliard
Georgia State University

Calvin Jones
Abt Associates

Adriana de Kanter
Office of Policy and Planning

Michael Kean
American Association of Publishers

Richard Maraschiello
The School District of Philadelphia

Phyllis McClure
NAACP Legal Defense Fund

William Padia
California Department of Education

Judy Pfannensteil
Research Training and Associates

Ted Rebarber
Office of Educational Research and Improvement

Edward Reidy
Kentucky Department of Education

Thomas Rosica
The School district of Philadelphia

Allen Schenck
RMC Research Corporation

Elois Scott
Office of Policy and Planning

Robert Slavin
The Johns Hopkins University

Thomas Timar
University of California, Riverside

Emily Wurtz
National Education Goals Panel

We would also like to acknowledge Nancy Kober for her skillful writing of this document.

Special thanks go to Department staff in the Office of Committee Management, Office of the General Counsel, Office of Policy and Planning, and the Office of Elementary and Secondary Education for their contribution to the Committee's operation.

Mary Jean LeTendre
Director
Compensatory Education Programs
U.S. Department of Education

ED/OPP93-22

BEST COPY AVAILABLE